

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 10:55:48 ; Search time 60 Seconds
(without alignments)
3569.517 Million cell updates/sec

Title: US-09-903-216-2

Perfect score: 4022

Sequence: 1 MAQRNNAKSSGNSSSSSGSS.....IVDVHPELTPQRRSLPAI 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1990s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	4022	100.0	758	4	AAB83919 A human a
2	4022	100.0	758	6	ABG72365 Human asp
3	4022	100.0	758	6	ADA00639 Human asp
4	1353.5	33.7	265	6	ABU92053 Human pro
5	1334	33.2	255	2	AAV33642 Human lab
6	1320	32.8	255	5	AAU85544 Clone #48
7	1320	32.8	255	6	ABU69516 Human lun
8	1320	32.8	255	6	ABU66419 Lung canc
9	933	23.2	422	4	ABBB1986 Drosophil
10	274.5	6.8	369	4	ABAB73682 Human oxi
11	271.5	6.8	236	3	AAAB43327 Human ORF
12	262	6.5	264	6	ABM68322 Photorhab
13	241	6.0	104	4	AAU29679 Novel hum
14	218.5	5.4	299	6	ABU21099 Protein e
15	217.5	5.4	324	4	AAU28081 Novel hum
16	215	5.3	47	4	AAU31979 Novel hum
17	204.5	5.1	3111	4	ABB60327 Drosophil
18	203.5	5.1	562	2	AAU70491 Leucocyto
19	199	4.9	109	4	AAU73672 Human oxi
20	199	4.9	783	2	AAU5804 C-termina
21	197.5	4.9	1018	2	AAU98747 P. vivax
22	197.5	4.9	1018	2	AAU97039 A. secreta
23	197.5	4.9	1018	4	AAU66528 Plasmodiu
24	197.5	4.9	1132	2	AAU97866 Chicken l
25	196.5	4.9	355	6	ABU23196 Protein e

26	191.5	4.8	1881	3	AAV44506 Streptoco
27	191.5	4.8	1881	6	ABU01047 S. pneumo
28	190.5	4.7	778	3	AAV46504 Arabidops
29	190.5	4.7	788	4	ABE68264 Drosophil
30	190.5	4.7	1162	3	AAV96255 Kaposi's
31	190.5	4.7	1162	3	AAV58500 HHV8 ORF
32	190.5	4.7	1162	4	AAV62331 Amino aci
33	190.5	4.7	1162	5	ABU05621 Kaposi's
34	190	4.7	360	2	AAW03627 Human fol
35	189.5	4.7	304	6	ABU17264 Protein e
36	189.5	4.7	312	6	ADA36728 Acinetoba
37	189.5	4.7	1616	6	ABU35669 Protein e
38	189.5	4.7	1616	7	ABO23515 Mycoplasma
39	189	4.7	3263	4	ABB67210 Drosophil
40	187.5	4.7	1192	4	ABBS9642 Drosophil
41	186.5	4.6	412	2	AAW03626 Human thy
42	186	4.6	49	6	ADA56732 Human sec
43	186	4.6	49	6	ADA40577 Human sec
44	186	4.6	49	6	ABR47638 Human sec
45	186	4.6	50	2	AAV01148 Secreted

ALIGNMENTS

RESULT 1
AAB83919
ID AAB83919 standard; protein; 758 AA.
XX
AC AAB83919;
XX
DT 23-JUL-2001 (first entry)
XX
DE A human aspartyl (asparaginyl) beta-hydroxylase (HAAH).
XX
KW Epidermal growth factor-like domain; EGF-like domain; cancer;
KW human aspartyl beta-hydroxylase; HAAH; malignant neoplasm; tumour.
XX
OS Homo sapiens.
XX
PN WQ200135102-A2.
XX
PD 17-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US030738.
XX
PR 08-NOV-1999; 99US-00436184.
XX
PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX
FI Wands JR, De La Monte SM, Ince N, Carlson RI;
XX
DR WPI; 2001-329171/34.
XX
DR N-PSDB; AAF69811.
XX
PT Diagnosing malignant neoplasm in a mammal, involves contacting mammalian
PT sample with antibody that binds to human aspartyl beta-hydroxylase
PT polypeptide to form antigen-antibody complex and detecting the complex.
XX
PS Disclosure; Page 5; 76pp; English.

XX
CC The present sequence represents a human aspartyl (asparaginyl) beta-hydroxylase (HAAH) enzyme. Epidermal growth factor (EGF)-like domains of polypeptides are hydroxylated by HAAH enzymes. HAAH is used in the method of the invention. The specification describes a method for diagnosing a malignant neoplasm in a mammal. The method comprises contacting a body fluid with an antibody which binds to HAAH polypeptide under complex forming conditions, and detecting the antigen-antibody complex. The method is useful for diagnosing and prognosing a malignant neoplasm in a body fluid e.g. central nervous system (CNS)-derived body fluid, blood, serum, urine, saliva, sputum, lung effusion, and ascites fluid of mammal, where the neoplasm is derived from endodermal tissue and is selected from colon cancer, breast cancer, pancreatic cancer, liver cancer, cancer of

Db 121 PEAEPHTPEEQVVEBAEPQNIIEDEAKEQIQSLHEMVAHAEVGEGLQEDGFTGP 180
 Qy QEDDFELMATDVRDRETELEVEVSHETSHYHVEETVSQDCNODMEEMSEQENPDSS 240
 Db QEDDFELMATDVRDRETELEVEVSHETSHYHVEETVSQDCNODMEEMSEQENPDSS 240
 Qy PVVEDERLHDDTDVTVQVVEAQVPLENEGIEITEVTAPDNPVEDSQVIVERVSI 300
 Db PVVEDERLHDDTDVTVQVVEAQVPLENEGIEITEVTAPDNPVEDSQVIVERVSI 300
 Qy FVVEEQVQVPEPTRKTDDEPKAKVKKPKLANKDKTKIKAEIDAAEKLKRGKIEEA 360
 Db FVVEEQVQVPEPTRKTDDEPKAKVKKPKLANKDKTKIKAEIDAAEKLKRGKIEEA 360
 Qy VNAFKELVRKYQPSPRARYGKAQCEDDLAEKRRNEVLRAETIYQEVASLPDVPADLLK 420
 Db VNAFKELVRKYQPSPRARYGKAQCEDDLAEKRRNEVLRAETIYQEVASLPDVPADLLK 420
 Qy LSLKRRSDRQOFLGHMRGSLTLQRLVQLFPNDTSLKNDLGCVGYLLIGDNDNAKKVYEEV 480
 Db LSLKRRSDRQOFLGHMRGSLTLQRLVQLFPNDTSLKNDLGCVGYLLIGDNDNAKKVYEEV 480
 Qy LSVTPNDGFAKHVGFILKAQNKIAESIPLYKEGIESGDPGTDGREFYHLGDAMQVRGN 540
 Db LSVTPNDGFAKHVGFILKAQNKIAESIPLYKEGIESGDPGTDGREFYHLGDAMQVRGN 540
 Qy KEAYKMYELGKHGRHAFASVQSRSLYVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
 Db KEAYKMYELGKHGRHAFASVQSRSLYVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
 Qy GLAVMDKAKGLFLPEDENLREKGDWSOFTLWQQRNENACKGAPKCTTLEKFPETGTC 660
 Db GLAVMDKAKGLFLPEDENLREKGDWSOFTLWQQRNENACKGAPKCTTLEKFPETGTC 660
 Qy RRGQIKYSIMHPGTHVPHGTPTNCRMLHGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
 Db RRGQIKYSIMHPGTHVPHGTPTNCRMLHGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
 Qy DSEFHEVQDASSPRLPIVDVWHPPELTPOQRSLPAI 758
 Db DSEFHEVQDASSPRLPIVDVWHPPELTPOQRSLPAI 758

RESULT 4

ABU92053
 ID ABU92053 standard; protein; 265 AA.

AC ABU92053;

DT 15-JUL-2003 (first entry)

XX Human protein modification and maintenance molecule-33 (PMMW-33).

XX Human; protein modification and maintenance molecule; PMMW; cancer;
 cell proliferation disorder; atherosclerosis; neurological disorder;
 epilepsy; Huntington's disease; stroke; immune disorder; allergy;
 inflammatory disorder; AIDS; developmental disorder; hypothyroidism;
 Cushing's syndrome; gastrointestinal disorder; epithelial disorder;
 infection; cytostatic; antiarteriosclerotic; anticonvulsant; nootropic;
 neuroprotective; cerebroprotective; anti-HIV; antiallergic; vulnerary;
 antiinflammatory; thyromimetic.

XX Homo sapiens.

XX WO2003031939-A2.

XX 17-APR-2003.

XX 11-OCT-2002; 2002MO-US032850.

XX 12-OCT-2001; 2001US-0329689P.

XX 25-OCT-2001; 2001US-0335703P.

XX 09-NOV-2001; 2001US-0348887P.

PR 28-NOV-2001; 2001US-0334145P.
 PR 06-DEC-2001; 2001US-0337451P.
 PR 14-DEC-2001; 2001US-0340584P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Rankumar J, Gorvad AE, Baughn MR, Emerling BM, Yang J, Lee SY;
 PI Tran UK, Becha SD, Duggan BM, Lee EA, Griffin JA, Li JX;
 PI Sprague WW, Hafalia AJA, Chawla NK, Lehr-Nason PM, Kable AE, Yue H;
 PI Marquis JP, Yao MG, Richardson TW, Tang TY, Jin P, Chien D;
 PI Bhacia U, Burrill JD, Lee S, Blake JJ, Ho A, Zheng W;
 XX WPI; 2003-430274/40.
 DR N-PSDB; ACA92448.
 XX New human protein modification and maintenance molecules (PMMW), useful
 PT for diagnosing, treating and preventing diseases or conditions associated
 PT with the aberrant PMMW expression e.g. cancer, atherosclerosis, or
 PT infections.
 XX Claim 1; Page 273-274; 311pp; English.

XX The present invention relates to the isolation of human protein
 CC modification and maintenance molecules (PMMW), and the polynucleotide
 CC sequences encoding them. A total of 40 PMMW polypeptides (designated PMMW
 CC -1 to PMMW-40) are disclosed. The sequences of the invention are useful
 CC for diagnosing a condition or disease associated with the expression of
 CC PMMW in a subject, preparing a polyclonal or monoclonal antibody, and
 CC generating an expression profile of a sample containing the
 CC polynucleotides. The diseases or conditions associated with decreased
 CC expression or overexpression of PMMW are cell proliferation disorders
 CC (e.g. cancer, atherosclerosis), neurological disorders (e.g. epilepsy,
 CC Huntington's disease, stroke), immune/inflammatory disorders, (e.g. AIDS,
 CC allergies), developmental disorders (e.g. hypothyroidism, Cushing's
 CC syndrome), gastrointestinal or epithelial disorders, and infections. The
 CC PMMW polypeptides or their fragments are useful in screening compounds
 CC for effectiveness as agonists or antagonists of the polypeptides, or in
 CC altering the expression of the target polynucleotide and compounds that
 CC specifically bind to, or modulate the activity of the polypeptide.
 CC ABU92021-ABU92060 represent the human PMMW polypeptides of the invention
 XX Sequence 265 AA;

Query Match 33.7%; Score 1353.5; DB 6; Length 265;
 Best Local Similarity 92.5%; Pred. No. 4.2e-99;
 Matches 259; Conservative 2; Mismatches 0; Indels 19; Gaps 1;

Qy 34 RETKHGKHNGRKGGLSGTSFFTFEMVIALGLGWTSAVWVFDLVYEEVLKGLIYDAD 93
 Db 5 KETKHGKHNGRKGGLSGTSFFTFEMVIALGLGWTSAVWVFDLVYEEVLKGLIYDAD 64
 Qy 94 GDGDFDVEDDAKVLGLKERTSEPAVPPPEAEPEHTEPEQVPEAEPPQNTDEAKEIQS 153
 Db 65 GDGDFDVEDDAKVLGLKERTSEPAVPPPEAEPEHTEPEQVPEAEPPQNTDEAKEIQS 124
 Qy 154 LLHEMVAHVEGEDLQEDGPTGEQDEDFELMATDVRDRETELEPEVSHETSHSYH 213
 Db 125 LLHEMVAHVEGEDLQEDGPTGEQDEDFELMATDVRDRETELEPEVSHETSHSYH 184
 Qy 214 VEETVSQDCNQDMEMSEQENEDSSSEFPVVEDERLHDDTDVTVQVVEAQVYEPLENEG 273
 Db 185 VEET-----DSSEFPVVEDERLHDDTDVTVQVVEAQVYEPLENEG 225
 Qy 274 IEITEVTAPDNPVEDSQVIVEEVSIFPVEEQQVEPPT 313
 Db 226 IEITEVTAPDNPVEDSQVIVEEVSIFPVEEQQVEPPT 265

RESULT 5

AAAY33642
 ID AAAY33642 standard; protein; 255 AA.
 XX AAAY33642;
 AC AAAY33642;

QY 179 PQEDDFLMATDVDDRFETLEPEVSHSETEHSYHVEETVSQDCNQDMEMMSQENPDS 238
 DB 121 PQEDDFLMATDVDDRFETLEPEVSHSETEHSYHVEETVSQDCNQDMEMMSQENPDS 180
 QY 239 SEPVEDERLHDDTDDVTYQVYEQAVYEPLENEGIEITEVTAPPENPVEDSQVIVEEV 298
 DB 181 SEPVEDERLHDDTDDVTYQVYEQAVYEPLENEGIEITEVTAPPENPVEDSQVIVEEV 240
 QY 299 SIFPVEEQQEVPPPT 313
 DB 241 SIFPVEEQQEVPPPT 255
 RESULT 7
 ABU69516
 ID ABU69516 standard; protein; 255 AA.
 XX
 AC ABU69516;
 XX
 XX
 DT 05-JUN-2003 (first entry)
 XX
 DE Human lung cancer-associated protein L979P.
 XX
 KW Human; lung cancer; lung tumour; cytostatic; vaccine; T cell expansion;
 KW CD4; CD8.
 XX
 OS Homo sapiens.
 XX
 XX US2002197669-A1.
 XX
 XX 26-DEC-2002.
 XX
 XX 03-MAY-2001; 2001US-00849626.
 XX
 XX 13-DEC-2000; 2000US-00736457.
 XX
 PA (BANG/) BANGUR C S.
 PA (FANG/) FANGER G R.
 PA (WANG/) WANG A.
 PA (WANG/) WANG T.
 PA (SWIT/) SWITZER A P.
 PA (MCNE/) MCNEILL P D.
 PA (CLAP/) CLAPPER J D.
 XX
 XX Bangur CS, Fanger GR, Wang A, Wang T, Switzer AP, Mcneill PD;
 PI Clapper JD;
 XX
 DR WPI; 2003-352750/33.
 DR N-PSDB; ACA12072.
 XX
 XX Novel lung cancer polynucleotide encoding lung cancer protein, useful for
 PT detecting the presence of lung cancer in a patient, and in pharmaceutical
 PT compositions, e.g. vaccines, for treating lung cancer.
 XX
 PS Example 5; Page; 72pp; English.
 XX
 CC The invention relates to a polynucleotide encoding a lung tumour protein,
 CC comprising a sequence selected from any of the 14 sequences mentioned in
 CC the specification, or a sequence (S2) mentioned in specification,
 CC complement of S1, sequences consisting of at least 20 contiguous residues
 CC of S1, sequences that hybridise to S1, sequences having 75%, preferably
 CC 90%, identity to S1, or degenerate variants of S1. Also included are an
 CC isolated polypeptide (comprising a sequence (S3) selected from any one of
 CC the 4 amino acid sequences mentioned in the specification, a sequence
 CC preferably 90%, identity to a sequence encoded by the polynucleotide), an
 CC expression vector comprising the polynucleotide operably linked to an
 CC expression control sequence, a host cell transformed or transfected with
 CC the vector, an isolated antibody (or its antigen-binding fragment) that
 CC specifically binds to the polypeptide, detecting the presence of a cancer
 CC in a patient, a fusion protein comprising the polypeptide, an
 CC oligonucleotide that hybridises to S1 under moderately stringent
 CC conditions, stimulating and/or expanding T cells specific for a tumour

CC protein (comprising contacting T cells with the polynucleotide, protein
 CC or antigen-presenting cells, under conditions and for a time sufficient
 CC to permit the stimulation and/or expansion of T cells) and inhibiting the
 CC development of a cancer in a patient (by incubating CD4⁺ and/or CD8⁺ T
 CC cells isolated from a patient with the polynucleotide, protein or antigen
 CC presenting cells that express the polynucleotide, such that T cells
 CC proliferate, administering to the patient an effective amount of the
 CC proliferated T cells, and thus inhibiting the development of a cancer in
 CC the patient. The polynucleotide, protein and cells are useful in a
 CC composition for stimulating an immune response in a patient, and for
 CC treating a cancer in a patient (particularly lung cancer). The
 CC oligonucleotide is useful for determining the presence of a cancer in a
 CC patient. The protein and oligonucleotides are useful in pharmaceutical
 CC compositions, e.g. vaccines. The polynucleotide is also useful as a probe
 CC or primer for nucleic acid hybridisation, and in the design and
 CC preparation of ribozyme molecules for inhibiting expression of tumour
 CC polypeptides and proteins in tumour cells. An amplified portion of the
 CC polynucleotide is useful for isolating a full-length gene from a suitable
 CC library. The present sequence is a protein encoded by a cDNA (full
 CC length, extended or partial) isolated from a library derived from lung
 CC tumour/ cancer cells. Note: The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from the USPTO at
 CC seqdata.uspto.gov/sequence.html?DocId=20020197669
 XX
 SQ Sequence 255 AA;

Query Match 32.8%; Score 1320; DB 6; Length 255;
 Best Local Similarity 98.8%; Pred. No. 1.9e-96;
 Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 59 MVIALLGWTSVAVVWFDLVYEEVLKGLGIYDADGDFDVKVLLGLKERSTSEPA 118
 DB 1 MVIALLGWTSVAVVWFDLVYEEVLKGLGIYDADGDFDVKVLLGLKERSTSEPA 60
 QY 119 VPPEAEPTPEEQVPPVPAEPQNIIDEAKEQIQSLHEMWHAEHVEGEDLQEDGPTGE 178
 DB 61 VPPEAEPTPEEQVPPVPAEPQNIIDEAKEQIQSLHEMWHAEHVEGEDLQEDGPTGE 120
 QY 179 PQEDDFLMATDVDDRFETLEPEVSHSETEHSYHVEETVSQDCNQDMEMMSQENPDS 238
 DB 121 PQEDDFLMATDVDDRFETLEPEVSHSETEHSYHVEETVSQDCNQDMEMMSQENPDS 180
 QY 239 SEPVEDERLHDDTDDVTYQVYEQAVYEPLENEGIEITEVTAPPENPVEDSQVIVEEV 298
 DB 181 SEPVEDERLHDDTDDVTYQVYEQAVYEPLENEGIEITEVTAPPENPVEDSQVIVEEV 240
 QY 299 SIFPVEEQQEVPPPT 313
 DB 241 SIFPVEEQQEVPPPT 255

RESULT 8
 ABU66419
 ID ABU66419 standard; protein; 255 AA.
 XX
 AC ABU66419;
 XX
 DT 22-MAY-2003 (first entry)
 XX
 DE Lung cancer therapy and diagnosis associated protein #43.
 XX
 KW Lung cancer; cytostatic; vaccine; gene therapy; cancer.
 XX
 OS Homo sapiens.
 XX
 XX US2002172952-A1.
 XX
 XX 21-NOV-2002.
 XX
 XX 10-JUL-2001; 2001US-00902941.
 XX
 XX 30-JUN-1999; 99US-00346492.

RESULT 10

AAAB73682
ID AAB73682 standard; protein; 369 AA.
XX
AC AAB73682;
XX
DT 11-SEP-2001 (first entry)
XX
DE Human oxidoreductase protein ORP-15.
XX
KW Human oxidoreductase protein; ORP; cell proliferative disorder;
KW arteriosclerosis; cirrhosis; psoriasis; cancer; endocrine disorder;
KW diabetes mellitus; diabetes insipidus; dwarfism; hirsutism; amenorrhea;
KW osteoporosis; metabolic disorder; obesity; phenylketonuria;
KW hypercholesterolaemia; reproductive disorder; infertility;
KW ovulatory defect; menstrual cycle defect; endometriosis; chromosome 22;
KW polycystic ovary disease; spermatogenesis disruption; impotence;
KW neurological disorder; epilepsy; stroke; Alzheimer's disease;
KW Huntington's disease; Parkinson's disease; Creutzfeldt-Jakob disease;
KW meningitis; cerebral palsy; muscular dystrophy; mood disorder; anxiety;
KW schizophrenic disorder; infection; autoimmune disorder;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS; asthma;
KW allergy; Crohn's disease; atopic dermatitis; gout; multiple sclerosis;
KW rheumatoid arthritis; ulcerative colitis; drug screening;
KW toxicity screening; transgenic animal; SNP detection; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WC200144448-A2.
XX
PD 21-JUN-2001.
XX
PF 07-DEC-2000; 2000WO-US033158.
XX
PR 16-DEC-1999; 99US-0172367P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Lal P, Tang YT, Hillman JL, Baughn MR, Azimzai Y, Lu DAM;
XX
DR WPI; 2001-390245/41.
DR N-PSDB; AAH24237.
XX
PT Novel human oxidoreductase protein (ORP) useful for diagnosing, treating
PT and preventing cell proliferative, neurological, viral, reproductive and
PT autoimmune/inflammatory disorders associated with abnormal expression of
PT ORP.
XX
PS Claim 1; Page 110-111; 136pp; English.
XX
CC Sequences AAB73668-AA73694 represent 27 novel human oxidoreductase
CC proteins, designated ORP-1 to ORP-27 respectively, and sequences AAH24223
CC -AAH24249 represent cDNAs encoding ORP-1 to ORP-27. Human ORP proteins
CC and nucleic acids are useful for diagnosing, treating or preventing cell
CC proliferative disorders (e.g. arteriosclerosis, cirrhosis, psoriasis,
CC cancers), endocrine disorders (e.g., type I or II diabetes mellitus,
CC diabetes insipidus, dwarfism, hirsutism, amenorrhea, osteoporosis);
CC metabolic disorders (e.g., obesity, phenylketonuria,
CC hypercholesterolaemia); reproductive disorders (e.g., infertility,
CC ovulatory and menstrual cycle defects, endometriosis, polycystic ovary
CC disease, disruption of spermatogenesis, impotence); neurological
CC disorders (e.g., epilepsy, stroke, Alzheimer's disease, Huntington's
CC disease, Parkinson's disease, meningitis, Creutzfeldt-Jakob disease,
CC cerebral palsy, muscular dystrophy, mood, anxiety and schizophrenic
CC disorders); viral, bacterial, fungal and parasitic infections; and
CC autoimmune/inflammatory disorders such as acquired immunodeficiency
CC syndrome (AIDS), allergies, asthma, Crohn's disease, atopic dermatitis,
CC gout, multiple sclerosis, rheumatoid arthritis or ulcerative colitis.
CC Human ORP proteins and nucleotides can be used to identify compounds
CC which modulate their activity or expression. ORP nucleic acid sequences
CC may also be used for assessing the toxicity of a test compound, to detect
CC upstream sequences such as promoters and regulatory elements, and to
CC create knock out or knock in animals or transgenic animals to model human

CC disease. Oligonucleotide primers derived from ORP gene sequences may be
CC used to detect single nucleotide polymorphisms (SNPs) and for mapping the
CC naturally occurring genomic sequences. Antibodies specific for ORP
CC proteins may be used in the diagnosis of disorders associated with
CC aberrant ORP expression, in assays to monitor patients being treated with
CC ORP or modulators thereof, and for assessing toxicity of potential drugs
XX
SQ Sequence 369 AA;
XX
Query Match 6.8%; Score 274.5; DB 4; Length 369;
Best Local Similarity 30.1%; Pred. NO. 5.4e-13;
Matches 75; Conservative 44; Mismatches 105; Indels 25; Gaps 9;
QY 526 RFYFLGDAMQVRGNKEAYKWEELG--HK-----RGHF-----ASVWORSLYNVNGLKAQPW 575
DB 125 KLYHNIQEVAKR-----YSWGMGRHKGIRGQRYLNSRPSIQKPEVFFDLPTTFY 178
QY 576 WTPKETYELVKSLERNWKLIRDEGLAVNDKAKGLFLPE--DENLRKGDWNSQFTLMQ 633
DB 179 FSRDAQKHD--VEVLERNFQTLICEPTLYKAPSNCSLPQGMKNSTPSCGEWTFYLVNQ 236
QY 634 GRNENACKGAPKCTCTLEKFPETTCGR-RGQIKYSIMHPGTHVWHTGTNCLRWHLG 692
DB 237 GUCVPRNCRKCPRTYRLGLSLRFTICGNVFNACISVLSFGTVITHEYGPTNIRIRCHLG 296
QY 693 LVIPKEGCKIRCANETRTWEEGKVLIFDDSFSEVHVQDASSF---RLIFTVDVWVHPELTP 749
DB 297 LKTP-NGCELVVGGEPQCAEGRCILFDDSFLLAAFHGSAEDGPRVFWVDLWHFNVA 355
QY 750 QQRSLPAI 758
DB 356 AERQALDFI 364
XX
RESULT 11
AAB43327
ID AAB43327 standard; protein; 236 AA.
XX
AC AAB43327;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF3091 polypeptide sequence SEQ ID NO:6182.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antiapoptotic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WC2000058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US008621.
XX
PR 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
PA (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;
 PI
 XX WPI; 2000-602362/57.
 DR N-PSDB; AAC77536.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 PS Claim 11; Page 5365; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiproliferative; antiparkinsonian; neurotrophic; neuroprotective; osteoparthy;
 CC anticonvulsant; antiarthritic; immunosuppressive; immunostimulant;
 CC cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
 XX
 SQ Sequence 236 AA;
 Query Match 6.8%; Score 271.5; DB 3; Length 236;
 Best Local Similarity 31.2%; Pred. No. 4.7e-13;
 Matches 72; Conservative 40; Mismatches 100; Indels 19; Gaps 8;
 QY 544 YKWEELG-HK-----RGHF-----ASVQWRLNVLNGLKAPWPTKTYTGLVKSLEARN 593
 DB 4 YSWGMRHKGIRGQRYLNSRPSIQKPEFFLPDPTTPYFSRDAQKH--VVVLERN 61
 QY 594 WKILRDEGLAVMDKAGLFLPE--DENLRKGDWSQFTLWQOGRNENACKGAPKTCILL 651
 DB 62 FQTLICEETLYKAFNSCLPQGWKNSTPSCGEMFTFLVNGVCVPRNCRKCPRTYRL 121
 QY 652 EKFPETTCGR-RGQIKYSIMHPGTHVWPHGTNCLRLMHLGLVTPKGGCKIRCANETRT 710
 DB 122 GSLRTICGNVFNACISVLSPGTVITEHYGPTNIRIRCHLGLKTP-NGCELVWVGEPQC 180
 QY 711 WEEGKVLIFDSDSFEHEVWDASSF---RLIFIVDVVHPELTQPQRRSLPAI 758
 DB 181 WAEGRCLLFDSDSLHAAPHEGSAEDGPRVVFVVDLWHPNVAARQALDFI 231
 RESULT 12
 ABM68322
 ID ABM68322 standard; protein; 264 AA.
 XX
 AC ABM68322;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Photorhabdus luminescens protein sequence #1419.
 XX
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 XX whooping cough.
 OS Photorhabdus luminescens.
 XX

PN WO200294867-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 07-FEB-2002; 2002WO-IB003040.
 XX
 PR 07-FEB-2001; 2001FR-00001659.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
 PI Buchrieser C;
 XX WPI; 2003-148459/14.
 XX
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX
 PS Claim 2; SEQ ID NO 1419; 1205pp; French.
 XX
 CC The invention relates to the isolation of genes and their encoded
 CC proteins from Photorhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of P. luminescens
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than P. luminescens and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by P.
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which P.
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated P. luminescens proteins
 XX
 SQ Sequence 264 AA;
 Query Match 6.5%; Score 262; DB 6; Length 264;
 Best Local Similarity 29.1%; Pred. No. 3.2e-12;
 Matches 66; Conservative 39; Mismatches 114; Indels 8; Gaps 3;
 QY 533 DAMQVGNKEAYKWYELGHKGHPASVWQSRSLYNVGLKAPFW-TPKBTGYTLVKSLE 591
 DB 35 DCLKEVED----MFYPLQQR---APLOADAKYIMPGLSTTFLWDTNFFPOLQPLVLSIM 87
 QY 592 RNWKLIRDEGLAVMDKAGLFLPEDENLRKGDWSQFTLWQOGRNENACKGAPKTCILL 651
 DB 88 NNADKIKOEINAVISGESQYITDYEHLGTCQDKWALYLFKNGQPNNAVANILPATWHIF 147
 QY 652 EKFPETTCGRGQIKYSIMHPGTHVWPHGTNCLRLMHLGLVTPKGGCKIRCANETRT 711
 DB 148 NNELRDWHCPLELVHFSVLQPGTVIKPHCDLWNFTNLHFAVDIPASCEIIVANEACRW 207
 QY 712 BEGKVLIFDSDSFEHEVWDASSFRLIFIVDVVHPELTQPQRRSLPAI 758
 DB 208 KEGRCLLFDSDSLHAAPHEGSAEDGPRVVFVVDLWHPNVAARQALVLI 254
 RESULT 13
 AAU29679
 ID AAU29679 standard; protein; 104 AA.
 XX
 AC AAU29679;
 XX
 DT 18-DEC-2001 (first entry)

```

XX Novel human secreted protein #170.
DE
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
XX WO200179449-A2.
PN
XX
XX 25-OCT-2001.
PD
XX
XX 16-APR-2001; 2001WO-US008656.
PF
XX
XX 18-APR-2000; 2000US-00552929.
PR
XX 26-JAN-2001; 2001US-00770160.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX
XX WPI; 2001-611725/70.
DR
XX
XX Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
PT
XX
XX Claim 20; Page 179; 765pp; English.
PS
XX
XX The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
XX Sequence 104 AA;
SQ
Query Match 6.0%; Score 241; DB 4; Length 104;
Best Local Similarity 54.7%; Pred. No. 3.7e-11;
Matches 47; Conservative 11; Mismatches 28; Indels 0; Gaps 0;
QY 13 SSSSGSGST$AGSSSPGARETKHGCHKGRGGLSGTSFPTFMVIALIGVWTSVAV 72
Dd 13 SAFTGTCSTSVRAAPGTPQPSMDAHTWGRKGRLSXTSFTTWSMVTALLGVWTSVSV 72
QY 73 VWFDLVYEEVLKIGIYDADGDGDF 98
Dd 73 VWFDLADYDDXIXALAIYDADGDVRF 98
RESULT 14
ABU21099
ID ABU21099 standard; protein; 299 AA.
XX
XX ABU21099;
AC
XX
XX 19-JUN-2003 (first entry)
DT
XX Protein encoded by Prokaryotic essential gene #6626.
DE
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
KW

```

```

XX Burkholderia fungorum.
OS
XX WO200277183-A2.
PN
XX
XX 03-OCT-2002.
PD
XX
XX 21-MAR-2002; 2002WO-US009107.
PF
XX
XX 21-MAR-2001; 2001US-00815242.
PR
XX 06-SEP-2001; 2001US-00948993.
PR
XX 25-OCT-2001; 2001US-0342923P.
PR
XX 08-FEB-2002; 2002US-00072851.
PR
XX 06-MAR-2002; 2002US-0362699P.
PR
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
PI
XX
XX WPI; 2003-029926/02.
DR
XX N-PSDB; ACA24969.
DR
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
PT
XX
XX Claim 25; SEQ ID NO 49023; 1766pp; English.
PS
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC on a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 299 AA;
SQ
Query Match 5.4%; Score 218.5; DB 6; Length 299;
Best Local Similarity 29.0%; Pred. No. 1.1e-09;
Matches 62; Conservative 29; Mismatches 96; Indels 27; Gaps 6;
QY 551 HKRGHFASVWQRSL-----YVNGLKAPQWMTPKETGVTLVKSLERWKL 596
Dd 17 HSRGKVRHGFRQLSDHSTFTAPLNGFVLFSAIPAQPYLPFSR--FPPEL-KLLKEEWRT 73
QY 597 IRDEGLAVMD----KAKGLFLPEDENLRKGDWSQFTLMQOGRNENACKGAPKTCITLLE 652

```

Db 74 IRDEAFALRDASHIRAAATAYNDIGNSPFRNGWRFFYLKWKYGRBPSAVALCPRTVELLG 133
Qy 653 KFPETTCGRGQIKYISIMHPTGTHVPHPTGPTNCLRMHLGLVLPK-EGCKIRCANEPTW 711
Db 134 RIPSVKAMFAQLP-----PGRLGLHRDPYAGALRYHLGLATPNHDGCAIWDGETYSW 188
Qy 712 BEGKVLFPDDSFHEVWQDASSFRLIFIVDVWHP 745
Db 189 RGDGDIVEDFVYLVAFNDTQEDRIILFCDIERP 222

RESULT 15
AAU28081
ID AAU28081 standard; protein; 324 AA.
AC AAU28081;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secretory protein, Seq ID No 250.
XX
KW Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;
KW gut protection; lung; liver fibrosis; immune deficiency; infection;
KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;
KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
KW fertility; analgesic; pain; antigen.
XX
OS Homo sapiens.
XX
PN WO200166689-A2.
XX
PD 13-SEP-2001.
XX
PF 05-MAR-2001; 2001WO-US004942.
XX
PR 07-MAR-2000; 2000US-00519705.
PR 19-MAY-2000; 2000US-00574454.
PR 17-JUN-2000; 2000US-00596193.
PR 14-JUL-2000; 2000US-00616847.
PR 19-SEP-2000; 2000US-00665363.
PR 20-OCT-2000; 2000US-00693267.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;
PI Zhao QA, Yang Y, Dzmanac RT, Zhang J, Chen R, Xue AJ, Wang J;
XX
DR WPI; 2001-589934/66.
DR N-PSDB; AAS44981.
XX
PT Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis and treatment of
PT cancer, neurological, inflammatory, and autoimmune disorders.
XX
PS Example 4; SEQ ID NO 250; 107pp; English.
XX
CC The invention relates to novel isolated human secreted polypeptides (I)
CC and polynucleotides (II). (I) and (II) are useful for treating
CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,
CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is
CC involved in increasing haematopoiesis, stem cell survival, bone growth
CC and remodeling. (I), (II) and modulators of (II) are useful for
CC prophylaxis or treatment of one or more cancers. (II) is also useful for
CC creating transgenic animals useful for studying the in vivo activities of
CC the polypeptide as well as for studying modulators of the polypeptides.
CC (I) induces the proliferation of neural cells and regeneration of nerve
CC and brain tissue and is useful for the treatment of central and
CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,
CC Parkinson's disease, Huntington's disease, and amyotrophic lateral

CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
CC and for regeneration of bone cartilage, tendon, ligament and/or nerve
CC tissue growth, and in tissue repair, healing of burns, incisions, ulcers,
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders,
CC or periodontal disease. Furthermore, (I) is also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, various immune deficiencies and
CC disorders including severe combined immunodeficiency (SCID), bacterial or
CC fungal infections, autoimmune disorders e.g. multiple sclerosis,
CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
CC reactions and conditions, such as asthma or other respiratory problems.
CC In addition, (I) affects biorhythms or circadian cycles of rhythms,
CC fertility, metabolism, catabolism, anabolism, storage or elimination of
CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
CC analgesic effects or other pain reducing effects, immunoglobulin like
CC activity and can act as an antigen in a vaccine composition to raise an
CC immune response. AAU28020-AAU28395 represent novel human secreted protein
CC amino acid sequences of the invention
XX
SQ Sequence 324 AA;
Query Match 5.4%; Score 217.5; DB 4; Length 324;
Best Local Similarity 36.6%; Pred. No. 1.5e-08;
Matches 53; Conservative 19; Mismatches 56; Indels 17; Gaps 5;
Qy 625 WS-----OFTLWQGR---RNENACKGAPKCTLLEKFPETTCRRGQIKYISIMHPG 673
Db 175 WSPFLAPGCYQLLYQAGRCQPSNCRCPGAYRALGLRSFMSAN--TFGNAGFSVLLPG 232
Qy 674 THVWPHPTGTCRLRMHLGLVLPKEGCKIRCANEPTWEEGKVLIFDDSFHEVWQDAS- 732
Db 233 ARLEGRCGPTNARVRCHLGLKIP-PGCELVVGGEPCQWAEHGCHLLVDDSLFHTVAHNSP 291
Qy 733 --SFLILFIVDVWHPDLTPQRRSL 755
Db 292 EDGPRWFIVDLWHPNVAGAEQAL 316

Search completed: May 5, 2004, 11:02:20
JOB time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 11:00:08 ; Search time 21 Seconds
(without alignments)
3472.056 Million cell updates/sec

Title: US-09-903-216-2

Perfect score: 4022

Sequence: 1 MAQRKNAKSSGSSSGSGS.....IVDVHHPDLTPQRRSLPAI 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3965.5	98.6	757	138423	aspartyl beta-hydr
2	3172	78.9	754	138423	peptide-aspartate
3	890	22.1	270	JC7792	cardiac junctate-1
4	798	19.8	872	T18861	probable peptide-a
5	263.5	6.6	186	T47148	hypothetical prote
6	216.5	5.4	1110	I51116	NF-180 - sea lamp
7	199	4.9	1271	A45555	glutamate rich pro
8	197.5	4.9	312	H83527	hypothetical prote
9	191.5	4.8	1881	H95076	zinc metalloprotei
10	189.5	4.7	1616	G64242	cycadherence-acces
11	188.5	4.7	792	T42963	hypothetical prote
12	187	4.6	706	A45990	junctional sarcopl
13	183	4.5	763	T08929	hypothetical prote
14	181	4.5	771	A33430	h-caldesmon - chic
15	179.5	4.5	411	S47436	flagellar antigen
16	178.5	4.4	729	S68191	triadin - human
17	177	4.4	301	A82601	aspartyl/asparagin
18	177	4.4	845	A45669	neurofilament trip
19	176	4.4	1094	S49313	protein kinase - s
20	174.5	4.3	1948	S00485	gene 11-1 protein
21	173.5	4.3	1871	D96796	probable heat shoc
22	172.5	4.3	1240	S52734	hypothetical prote
23	172	4.3	465	S29796	hypothetical prote
24	172	4.3	630	S29796	hypothetical prote
25	171.5	4.3	607	S27776	80K protein (allel
26	171	4.3	1804	T34518	nestin - golden ha
27	169.5	4.2	592	B48315	lamin B2 - mouse
28	169.5	4.2	1233	S56271	hypothetical prote
29	168.5	4.2	695	T40168	hypothetical prote

30	168.5	4.2	913	2	T52485	neurofilament prot
31	168	4.2	302	2	AF1021	probable membrane-
32	168	4.2	1192	2	A71623	probable secreted
33	167.5	4.2	1876	2	E97944	zinc metalloprotei
34	166.5	4.1	1624	2	T25592	hypothetical prote
35	165	4.1	651	2	S18874	nucleolin - Africa
36	164.5	4.1	7962	2	I38346	elastic titin - hu
37	164	4.1	1957	2	T38077	hypothetical coile
38	163.5	4.1	1132	2	T43483	translation initia
39	163	4.1	411	2	S48647	peptidylprolyl iso
40	163	4.1	501	2	C71948	hypothetical prote
41	163	4.1	1877	2	T21861	hypothetical prote
42	163	4.1	1999	1	S21801	myosin heavy chain
43	162.5	4.0	299	2	B81081	hypothetical prote
44	162.5	4.0	793	1	JH0628	caldesmon - human
45	162.5	4.0	992	2	T46337	hypothetical prote

ALIGNMENTS

RESULT 1

I38423

aspartyl beta-hydroxylase - human

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jan-2000

C:Accession: I38423

R:Korioth, F.; Gieffers, C.; Frey, J.

Gene 150, 395-399, 1994

A:Title: Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase

A:Reference number: I38423; MUID:95121937; PMID:7821814

A:Accession: I38423

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-757 <RES>

A:Cross-references: EMBL:U03109; NID:9458031; PIDN:AAA2108.1; PID:9458032

C:Superfamily: peptide-aspartate beta-dioxygenase; tetratricopeptide repeat homology

F:54-75/Domain: transmembrane #status predicted <TRM>

F:341-374/Domain: tetratricopeptide repeat homology <TTR>

Query Match	98.6%;	Score	3966.5;	DB	2;	Length	757;		
Best Local Similarity	99.2%;	Pred. No.	1e-195;						
Matches	752;	Conservative	2;	Mismatches	3;	Indels	1;	Gaps	1;
QY	1	MAQRKNAKSSGSSSGSGSTAGSSSPGARRETKHGKHNGRKGGLSGTSPFTWFMV	60						
Db	1	MAQRKNAKSSGSSSGSGSTAGSSSPGARRETKHGKHNGRKGGLSGTSPFTWFMV	60						
QY	61	IALLGWTSVAVVWFVLDVYEEVLKGLGIYDADGDGDFVDKAVLLGLKERSTSEPAVP	120						
Db	61	IALLGWTSVAVVWFVLDVYEEVLKGLGIYDADGDGDFVDKAVLLGLKERSTSEPAVP	120						
QY	121	PEEAHPTEPEEQVPVEAPQNIIEAKEQIQSLHMHVHAHVEGEDLOQEDGPTGPQP	180						
Db	121	PEEAHPTEPEEQVPVEAPQNIIEAKEQIQSLHMHVHAHVEGEDLOQEDGPTGPQP	180						
QY	181	QEDDEFMATVDVDRFETLEPEVSHBETHSHYVEETVQDCNQDMEMMSQENPDOSSE	240						
Db	181	QEDDEFMATVDVDRFETLEPEVSHBETHSHYVEETVQDCNQDMEMMSQENPDOSSE	240						
QY	241	PVVEDERLHHTDDVTQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVS	300						
Db	241	PVVEDERLHHTDDVTQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVS	300						
QY	301	FPVEEQEVPPTNRKTDDPEQAKVKKKKPKLNKFKDTIKAEALDAAEKLRKRGKIEEA	360						
Db	301	FPVEEQEVPPTNRKTDDPEQAKVKKKKPKLNKFKDTIKAEALDAAEKLRKRGKIEEA	360						
QY	361	VNAFKELVRKYPOSPPARYGKAQCEDDLAEKRRSRNEVLRGAIETYQEVASLPDVPADLLK	420						
Db	361	VNAFKELVRKYPOSPPARYGKAQCEDDLAEKRRSRNEVLRGAIETYQEVASLPDVPADLLK	420						
QY	421	LSLKRSDRQOFLGHRMGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVVEEV	480						

Db 421 LSLKRRSDRQOFLHMGSLTLTLQLVQLFENDTSLKNDLGVGLLIGDNDNAKYYEEV 480
 QY 481 LSVTPNDGFAKHYGFILKAQNKIAESIPLYKIEGSDGDTDDGRFYFHLGDAMQVRGN 540
 Db 481 LSVTPNDGFAKHYGFILKAQNKIAESIPLYKIEGSDGDTDDGRFYFHLGDAMQVRGN 540
 QY 541 KEAYKYELGKRGHGFASVWQSLYNNVGLKAQSWTPKGTGYTELKSLERNWKLIRDE 600
 Db 541 KEAYKYELGKRGHGFASVWQSLYNNVGLKAQSWTPKGTGYTELKSLERNWKLIRDE 599
 QY 601 GLAVMDKAKGLFLPEDENLREKGDWQFTLWQOGRNENACKGAPKCTTLLKFPETTCG 660
 Db 600 GLAVMDKAKGLFLPEDENLREKGDWQFTLWQOGRNENACKGAPKCTTLLKFPETTCG 659
 QY 661 RRGQIKYSIWHGPHVHPHTGPTNCRMLHGLVPIKKGKIRCANETRWEGSKVLIFD 720
 Db 660 RRGQIKYSIWHGPHVHPHTGPTNCRMLHGLVPIKKGKIRCANETRWEGSKVLIFD 719
 QY 721 DSFEHEVQDASSFRLLIFIVDVHPELTPOQRSLPAI 758
 Db 720 DSFEHEVQDASSFRLLIFIVDVHPELTPOQRSLPAI 757

RESULT 2

BABOH

N:Altenate names: aspartyl (asparaginyl) beta-hydroxylase
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 31-Dec-1993 #sequence_revision 10-Feb-1995 #text_change 11-Jun-1999
 R:Jia, S.; Vandusen, W.J.; Diehl, R.E.; Kohl, N.E.; Dixon, R.A.; Elliston, K.O.; Stern, J. Biol. Chem. 267, 14322-14327, 1992
 A:Title: cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.
 A:Reference number: A42969; MUID:92332546; PMID:1378441
 A:Accession: A42969
 A:Molecule type: mRNA
 A:Residues: 1-754 <JIA>
 A:Cross-references: EMBL:M91213; NID:g162693; PID:AAA03563.1; PID:g162694
 A:Experimental source: brain
 A:Note: sequence extracted from NCBI backbone (NCBIP:108534)
 R:Wang, Q.; Vandusen, W.J.; Petroski, C.J.; Garbky, V.M.; Stern, A.M.; Friedman, P.A. J. Biol. Chem. 266, 14004-14010, 1991
 A:Title: Bovine liver aspartyl beta-hydroxylase. Purification and characterization.
 A:Reference number: A39470; MUID:91310689; PMID:1856229
 A:Accession: A39470
 A:Molecule type: protein
 A:Residues: 289-328 <WANG>
 A:Accession: B39470
 A:Molecule type: protein
 A:Residues: 615,'X',617-630,'XX',633-634,'X',636,'XX',639-641 <WA2>
 A:Accession: C39470
 A:Molecule type: protein
 A:Residues: 311-347,'X',349,'X',351-373,'X',375-379,'X',381-382 <WA3>
 C:Comment: This enzyme uses ferrous iron as a cofactor, and while beta-hydroxylating the C:Comment: Aspartic acid and asparagine residues in the EGF homology domain of certain F C:Superfamily: peptide-aspartate beta-dioxygenase; tetratricopeptide repeat homology C:Keywords: glycoprotein; oxidoreductase; transmembrane protein
 F:2-56/Domain: intracellular #status predicted <INC>
 F:57-78/Domain: transmembrane #status predicted <TRM>
 F:289-754/Product: peptide-aspartate beta-dioxygenase, 56K form #status predicted <56K>
 F:311-754/Product: peptide-aspartate beta-dioxygenase, 52K form #status predicted <52K>
 F:337-370/Domain: tetratricopeptide repeat homology <TT1>
 F:371-404/Domain: tetratricopeptide repeat homology <TT2>
 F:13,96,466,702/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 78.9%; Score 31.72; DB 1; Length 754;
 Best Local Similarity 78.7%; Pred. No. 4.8e-155;
 Matches 612; Conservative 45; Mismatches 77; Indels 44; Gaps 6;
 QY 1 MAQRKNAKSSG--NSSSSGSGSGSTAGSSSPGARRETKHGKNGKGLSGTSPFTW 57
 Db 1 NAPIRKAQGGGNSSSSGSGTCTGCTGSSSPGARRETKHGKNGKGLSGSGSFTW 60

QY 58 FMVIALLGWTSVAVVWFDLVDYEEVL-----GKLGIDYDADGDGDFDVD 102
 Db 61 FMVIALLGWTSVAVVWFDLVDYEEVLAKAKAFRYMLSEVLQKGLGIYDADGDGDFDVD 120
 QY 103 AKVLLGLKERSSTSEPAVPEEAPHTPEPEEQVVEAPQNIEDAEKEQIQLSLHEVHAE 162
 Db 121 AKVLLGLKEKPAKPTVPPEADMYPLWEDQVLESFGRQNIEDVVEVQVS-LDETVYSE 179
 QY 163 HVEGDLQOBDGDTGTPQOEDDEFMATVDVDDRFETLEPEVSHHEETSHSVHETVSDC 222
 Db 180 --PGENLPQEPGAELQPDHVFVGSADDDRYEPMTGTGAHHEETSDSYHIBETASPAY 237
 QY 223 NDMEMWSEQENPDSSSEPVED--ERLHHDITDDVTVQVVEQAVVEPLENEGIEITEVT 280
 Db 238 SQMEDMYEQENPDSSSEPVED--ERLHHDITDDVTVQVVEQAVVEPLENEGIEITEVT 280
 QY 281 APEDNPVEDSVIVBEVSIFPVVEEQEVPEPNRKTDDPEQAKVKKKPKLLNKPKDT 340
 Db 281 ---DHAVONSNTLSEPHMPPAEQEVPEPNRKTDDPEQAKVKKKPKLLNKPKDT 336
 QY 341 IKAELDAAEKLRRGKIEEAVNAFKELVRYKPOSPRARYKCAQEDDLAEKRRSNEVLRG 400
 Db 337 IKAELDAAEKLRRGKIEEAVNAFKELVRYKPOSPRARYKCAQEDDLAEKRRSNEILRR 396
 QY 401 AIETYQEVASLPDVPADLLKSLKRRSDROOFLGHMRGSLTLTLQLVQLFPNDTSLKNDL 460
 Db 397 AIETYQEVASLPDVPADLLKSLKRRSDROOFLGHMRGSLTLTLQLVQLFPNDTSLKNDL 456
 QY 461 GVGYLIGDNDNAKYYEVLSTPNDGFAKHYGFILKAQNKIAESIPLYKIEGSDGDP 520
 Db 457 GVGYLIGDNDNAKYYEVLSTPNDGFAKHYGFILKAQNKIAESIPLYKIEGSDGDP 516
 QY 521 GTDDGRFYFHLGDAMORVGNKEAYKMYELGKRGHGFASVWQSLYNNVGLKAQSWTPKE 580
 Db 517 GTDDGRFYFHLGDAMORVGNKEAYKMYELGKRGHGFASVWQSLYNNVGLKAQSWTPKE 576
 QY 581 TGYTELKSLERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWQFTLWQOGRNENA 640
 Db 577 TGYTELKSLERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWQFTLWQOGRNENA 636
 QY 641 CKGAPKCTTLLKFPETTCGRRQIKYSIMHPGTHVHPHTGPTNCRMLHGLVPIKKG 700
 Db 637 CKGAPKCTTLLKFPETTCGRRQIKYSIMHPGTHVHPHTGPTNCRMLHGLVPIKKG 696
 QY 701 KIRCANETRWEGSKVLIFDSSFEHEVQDASSFRLLIFIVDVHPELTPOQRSLPAI 758
 Db 697 KIRCANETRWEGSKVLIFDSSFEHEVQDASSFRLLIFIVDVHPELTPOQRSLPAI 754

RESULT 3

JC7792

cardiac junctate-1 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 02-Apr-2002 #sequence_revision 02-Apr-2002 #text_change 03-May-2002

C:Accession: JC7792

R:Hong, C.S.; Kwak, Y.G.; Ji, J.H.; Chae, S.W.; Kim, D.H.

Biochem. Biophys. Res. Commun. 289, 882-887, 2001

A:Title: Molecular cloning and characterization of mouse cardiac junctate isoforms.

A:Reference number: JC7792

A:Contents: Heart

A:Accession: JC7792

A:Molecule type: mRNA

A:Residues: 1-270 <HON>

A:Cross-references: GB:AF302653

C:Comment: This protein, a Ca2+ binding protein, plays a role both in contractile and co

C:Keywords: cardiac muscle; heart

Query Match 22.1%; Score 890; DB 2; Length 270;

Best Local Similarity 63.9%; Pred. No. 8.8e-39;

Matches 179; Conservative 30; Mismatches 57; Indels 14; Gaps 5;

QY 34 RETKHGKNGRKGSLSGTSTFTWVIALLGWTSVAVVWFDLVDYEEVLGKGIYDAD 93

Db 5 KEAKGGHKGRRGSGSFFTFWFWALLGVWTSVAVVWFDLVYEEVLGKLVGDAD 64
Qy 94 GDGDFVDDAKVLGLKERTSEPAVPEEAPTEPEEQVPEEAPQNTIDEAKQIQS 153
Db 65 GDGDFVDDAKVLGLKERTSPERTFP- EAETHAELEEQAPGADIQNVEDEVKEIQS 123
Qy 154 LHEMVHAEHVEGDLQOEGPTCEQEDDERFLMATDVEDRFTLPEVSHETESYH 213
Db 124 LLQESVHTDR----DL-EADGLAGEPOVEVDFLTVTDSDDRFDELPGVHBEIETYH 178
Qy 214 VEETVSQDCNQDEEMMSQENPSSBFFVEDERLHDTDDVTYQVVEEQAVYPLENEG 273
Db 179 VEDTASQHPNDMEEMTNEQNSDPSEAVTDAGVLLPHAEVRHODYE-PVYEPSHEG 237
Qy 274 IEITEVTAPEDNPEVDSQVIVVEVSIFPVEEQEVPEP 313
Db 238 VEIS-----DNTIDDSIISEINVASVEEQQTTPDT 270

RESULT 4
T18861
probable peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18861; T23533
R:Swinsburne, J.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19032
A:Accession: T18861
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-872 <W1>
A:Cross-references: EMBL:Z79596; NID:el323798; PIDN:CA801859.1; GSPDB:GN00028; CESP:K09A
A:Experimental source: clone C02C6
R:Swinsburne, J.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19753
A:Accession: T23533
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-872 <W12>
A:Cross-references: EMBL:Z79601; PIDN:CA801887.1; GSPDB:GN00028; CESP:K09A9.6
A:Experimental source: clone K09A9
C:Genetics:
A:Gene: CESP:K09A9.6
A:Map position: X
A:Introns: 18/1; 40/3; 74/1; 104/2; 139/1; 170/3; 385/3; 432/2; 464/3; 530/2; 609/3; 776
C:Keywords: oxidoreductase

Query Match 19.8%; Score 798; DB 2; Length 872;
Best Local Similarity 26.8%; Pred. No. 2e-33;
Matches 191; Conservative 144; Mismatches 278; Indels 100; Gaps 15;

Qy 89 TYDADGDFDVFDDAKVLGLKERTSEPAVPEEAPTEPEEQ-----VPVEAPQ 141
Db 205 VEDDDDDDDDDDDVE-----APAPQESRQKAKHOREKKKKNKYQPVKEEPD 255
Qy 142 NIEDEAKQIQSLHEMVHAEHVEGEDLQOEG-----PTGEPOQEDDEFLMATDVDR 195
Db 256 LDD 315
Qy 196 FETL-----EPEVSHETE--HSYHVERT-----VSQDC-NODMEE- 228
Db 316 SDKVEEKNDGDEDFHVSUREAQRRLRQLHRSRDRNRPRQGNRECHQDCPNRESLXP 375
Qy 229 ----MMSEQNPDSPEVVEDERLHDTDDVTYQVVEEQAVYPLENEGIEITEVTAPPE 284
Db 376 RKSLLVTKKTKSMVERILLDFDDEDDDD-----EDDDSEVKN-----N 415
Qy 285 DNPVEDSQVIVEVSIFPVEEQEVPEPTEPNRKTDDDEQAKVKKKPKLLNFKDXTIKAE 344
Db 416 DREDDDDDDVDERIS-----DRDSSSYKRRHAITTEEIGFRDI 454

Qy 345 LDAAEKLRKGKIEEAVNAFKELVKRYQSPRARYGKAQCEDDDLAEKRRSNEVLRAIET 404
Db 455 LDRADNLVEKHQVEEAMELFDHVIAYVPASTRAYFGKARAYDINGEATEDETDRDKALEI 514
Qy 405 QOEVASLPYDPADLLKLKRRSDROOFLGMRGSLTLTQLRVLQFPNDTSLKNDLVGY 464
Db 515 YEKILQNSGVFDALFRQAQRLTEKTRFGRQLKHTLTAHRYFIDRFPPEELNLOTDFAFS 574
Qy 465 LLTGDDNDAKYYEEVLSVTPNDGFAKVHYGFILKA-QNKIAESIPYLKGGIESGDPCTD 523
Db 575 VMKRYEDARTVLKLVLANDPNHVIAYYILKAHDDKVEQGVVALMRKSLKNADNEIT 634
Qy 524 DGRFYFHLGDAMORVGNK-EAYKMYELGHRGHFASVWQRSLYVNVGLKQAPWMTPKETG 582
Db 635 DPKFYQLGHGLITTLGRKSEADAVYQKAAQGVMTAQQRSLYNIIEGLTGRAMWMDQTP 694
Qy 583 YTELKSLERNKWLIRDEGLAVMDKAKGLFPEDEENREKGDWSQFTLWOGRNRNACK 642
Db 695 YSKFLKTVERQWATIROEGMEVLKDCSDCLMDHNOQLVIDGQMKFFPTMSEQNFKSCE 754
Qy 643 GAPKTCITLLEKFPETTCRRGQIKYSIMHPGTHVWPHPTNCRLRMHLGLVPIKEGCKI 702
Db 755 RMPQTCILIOEFAASSNASKDMHLSVSSGASTLPHCGPTNYHLQAHGLVSPSE-ARI 813
Qy 703 RCANETRTWEEGKVLIFDDSPFEHEVMOD--ASSFRILFIVDVWHPBLTPOOR 752
Db 814 RVGNETKGRSGKFIYYDDSFHEHLOFDGASSSSFRVLVITLQLMHPEVQPHOR 866

RESULT 5
T47148
hypothetical protein DKFZp761P039.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47148
A:Reference number: Z24379
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A:Accession: T47148
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-186 <AAA>
A:Cross-references: EMBL:AL161993
A:Experimental source: adult amygdala; clone DKFZp761P039
C:Genetics:
A:Note: DKFZp761P039.1

Query Match 6.6%; Score 263.5; DB 2; Length 186;
Best Local Similarity 35.4%; Pred. No. 6.5e-07;
Matches 63; Conservative 28; Mismatches 80; Indels 7; Gaps 4;

Qy 587 VKSLERNWKLIRDEGLAVMDKAKGLFLE--DENLREKGDWSQFTLWOGRNRNACKGA 644
Db 5 VEVLERNFOTTLCEFTETLYKAFNSCSLPQGNKMSPTSGSEWFTFVLVNGQVCPNCRKC 64
Qy 645 PKTCTLLEKFPETTCGR-RGQIKYSIMHPGTHVWPHPTNCRLRMHLGLVPIKEGCKIR 703
Db 65 PRTYLLGSLRTICGNVYFVGNACISVLSPGVITEHYGPTNIRIRCHLGLKATP-NGCELV 123
Qy 704 CANETRTWEEGKVLIFDDSPFEHEVMQDASSF---RLIFIVDVWHPBLTPOORRSLPAI 758
Db 124 VGGEPQCAEGRCILLFDDSLFAAFHEGSABDGPRVFMVLDLWHPNVAARQALDFI 181

RESULT 6
I51116
NF-180 - sea lamprey
C:Species: Petromyzon marinus (sea lamprey)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: I51116
R:Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995

QY 702 IRCANETRTWEGKVLIFDDSPHEVWODASSFRILFIVDWHMP 745
Db 185 IYVGGQPYAWRGEDVMEFTFVHVWKNKTEOTRVLFCDIERP 228

RESULT 9
H95076
zinc metalloproteinase ZmpB, probable [imported] - Streptococcus pneumoniae (strain TIGR
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95076
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95076
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1881 <KUR>
A:Cross-references: GB:AE005672; PID:AAK74809.1; PID:g14972138; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0664

Query Match 4.8%; Score 191.5; DB 2; Length 1881;
Best Local Similarity 21.6%; Pred. No. 0.06;
Matches 138; Conservative 82; Mismatches 226; Indels 193; Gaps 30;

QY 111 ERSTSEPAVPEE--AEPHTEPEEQVPEAEQNTDEAKEIQSLHMHVHAHVHVED 168
Db 211 KEDSAPPA-PVEVGEVSEKPEKAVAPESQPSDKPAESKVEQAGEPV-----APRED 265

QY 169 LQEGDGTGTEPOQEDDEFMLATDVDDRFETLPEVSHETSHTVSVDCNQ-DME 227
Db 266 EKAPVEPEKQPEAPEE--KAVEETPKQESTPDTKAEETVEP--KEETVNSIQPKVE 321

QY 228 ENMSQENPDSPEVVEDE----RLHHTDDVTYQVYEQAVPEPLENEGIELITEVAPP 283
Db 322 TPAVEKQTEPTPEKVEQAGEVAPREDQAPTAPVEPEKEVPEEERAV---EETPKP 378

QY 284 ED-----NPVEDSQV---IVEEVSIRPVE-----EQO 307
Db 379 EDKIGKIGTKEPVDKSELNNQDKASSVPTDYSTRASYNALGPVLETAKGVASEPVKQP 438

QY 308 EYVPEPTR-----KTD-----DPEQKAKYKKKPKLLNKFDKTIKAEIDAAEK 350
Db 439 EVNSETNKLKTAIDALNVSKTELNNITADAKTKVKEHYSRQWNLQTEVTKAEKVAANT 498

QY 351 LKRGKIEAVNAF-----KELVVKYQSPSPARYGKQ 383
Db 499 DAQOSEWNAEKLKATIEKLVELSEKPILTLTSTDKILEREAVAKYTL-----NQMKTK 555

QY 384 CEDDLAEKRSNEVLKGA-----ETVQ-----E 407
Db 556 IKSITAEIKKGEVINTVLTDDKVTETISAFKNLEYKEYTLSTTMYDNGGEETE 615

QY 408 VASLPDVPADLLKLSLK--RRSDROQFLGHMRGSLTLQRLVOLFPNDTSLKNDLGVGYL 465
Db 616 TLENQNIQLDLKVELKNKRTDLIKY-----ENKSTNESLITTIPODKSNY-----YL 665

QY 466 LIGDNDN-----AKVYEVLSVTNDGFAKHYGFILK-AQNKTAES-IPYLKEGIES 517
Db 666 KITSNNQKTTLLAVKNIEETVNGTTPYKVTAADNLVSRADNKPFEYVHYIEK-----721

QY 518 GDPDGTDRGFYFLHGDAMORVGNKEAYKYVELGHKRGHSPVWQRSLXNV--NGLKAQPW 575
Db 722 --PKVHEDVNYNFKELVEAIQN-DPSKEYRLQO-----SMSARNVVPNG---KSY 766

QY 576 WTPKETGYTELKSLERNKWLIRDEG--LAVMDKAKGLF 612

Db 767 ITKEFTG-----KLLSSEKQFAITELEHPLF 793

RESULT 10
G64242
cytadherence-accessory protein (hmw1) homolog MG386 - Mycoplasma genitalium
C:Species: Mycoplasma genitalium
C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 07-Dec-1999
C:Accession: G64242
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhmann, J.; Nguyen, D.; Utterback, T.R.; Saudak, D.M.; Phillips, C.A.; Merrick, J.A.
; C.A.; Venter, J.C.
Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346; PMID:7569993
A:Accession: G64242
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1616 <TIGR>
A:Cross-references: GB:U39723; GB:L43967; NID:g1046092; PID:g1046097; TIGR:MG386
A:Experimental source: strain G-37
C:Genetics:
A:Genetic code: SGC3

Query Match 4.7%; Score 189.5; DB 2; Length 1616;
Best Local Similarity 22.3%; Pred. No. 0.063;
Matches 142; Conservative 78; Mismatches 205; Indels 213; Gaps 35;

QY 75 FDLVDYEEVLGK-----LGIYD-----ADGDGDFVDVDAKVL-----106
Db 1091 FDTVKHEAVFDKNQQTGTELEPEQVSEAEVVDQTDTTVGSEPAVFDVQPEKTTVEKFD 1150

QY 107 -LGLKERSTSEPAVPEEAEHTEPEEQVPEAEQNTDEAKEIQSLHMHVHAHVHVE 165
Db 1151 DVENQOKVISEPQVQCEAVPEPSAKAFDPSVESQDSQPEVLEEVQTPQLQVPE 1210

QY 166 GEDLQEQDGTGTEPOQEDDEFMLATDVDDRFETL-PEVSHETSHTVSVDCNQ-CN 223
Db 1211 SQPEATFD--TVQPTQPE---AKFDSPTVETVQPESESSEPTQO--HVESEASFDEEN 1262

QY 224 QMEMMSEQENP--DS-----SEPVVERLHHTDDVTYQV-----YEQAVYEPL 269
Db 1263 YDFDEPNYDFDQPSYSDLSQLPSEPDYVDENYDFDEPNYETESKESBEPQEPVQEQP- 1321

QY 270 ENEGIELITEVAPPE-DNPVE--DSQ--VIVEEVSIRPVEEQVPEPPTNKTDDPEOK 323
Db 1322 ---GEAVPEPSAKAFDPSVESQDSQPEPLEEEVQTQPEIQFVSESQPEATFDTVQPEQT 1378

QY 324 AKVKKKKPKLLNKFDKTIKAEIDAAEKLRKRGKIBEAVNAFKELVRKYQSPSPARYGKQ 383
Db 1379 PQEA-----KFD-----SP-----1387

QY 384 CEDDLAEKRSNEVLKGAETVQPE--VASLPDVPADLLKLSKRRSDROQFLGHMRGSL 441
Db 1388 -----VETIQEPQVSSEPEV--VVPQNFEEPKPE-----TVL 1417

QY 442 TLQRLVOLFP---NDSLTKNDLGVGYLLIGDNDNAKVVYEVLSVTPNDGFAKHYGFIL 498
Db 1418 EEPQADEIQPEASEBESLWELLVG-----NNSYGHYEP-----DG-EWVWAGFFG 1462

QY 499 KAQ--NKTAESIPYLKEGIESGDPGTDGGRFYFHL-GDAMQ-RVGNKEAYKYVELGHKRG 554
Db 1463 DQKWNKA-TVKWARE-----RDYLPILIGDEVYGRYNNKGEMWYGFYDESG 1509

QY 555 HFASV---WQSLYVNGLKAQFWTP---KETGYTELKSLERNKWLIRDEGLAVMDK 607
Db 1510 DWLVYDEQWKNRQPRIN--EAPKFEKLIQNGEYGYE---DNEWNW-----1551

QY 608 AKGLFLIPEDENLRKGDWSQFTLWQQGRENENACKGAP 645
Db 1552 -----YDGEFDSGNNWLVFQSEETENLNEDITDIP 1582

QY	161	AHVEG--EDLQQDDGGTGEQQDD-----EFLMATDV-----DORFETLEPEVS	204
Db	92	VTKDEGQAENMDEDADGKKEQTDDGVSDVTMKENVESKNNAKYAKDEXKETKTDT	151
QY	205	HEETHSVHVETVTSQCNCQ-----DMEE--MMSEQENPDSBPVVVEDERLHH	251
Db	152	--EADHKAGKHEDIHEADKANGTKDGNTGDJKEEGTLVDDEKGTDM-----DEKVENG	203
QY	252	TDDVTYQVYESQAVYEPLNKGIEITETVAPPDENPVDSQVIVEVS-----IPFPVEEQ	307
Db	204	DENKOVENVGEKEDEBENTKEVEAAKAEBVDESKEVDEKEGEDENDNEKVESDAKE	263
QY	308	EVPPTNKTTDDPEOKAVKKVKPKLLNKFDKTIKAELDAAEKLRKGKTEEAVNFAKEL	367
Db	264	DEKETNDDEKDEESKSGKRK-----GTSSGKVRKKNKTEEV-----	305
QY	368	VKYPQSFRARYGKAQCEDDLAEKRRRNEVLRGAI--ETYOEV-----ASLPDVPADL	418
Db	306	--KDAEPRTPF-----SDRPVRERKSVERVALUIDKDSSKEFVBKRGAYLKDPNVA	358
QY	419	LXLSLKRSRDQQL-----GHMGSLLTQLRVQLFPNDTSLKNDLG-V-GYLIGDGN	472
Db	359	NKVMKRSDETLLHPILFGRRGKAQIK-----TNILGFSGFWHGDEKK	406
QY	473	AKKVEEVLSTPNPD---GFARKHYGFILKAQNKIAESIPLYKEGIE---SGDPCTDGO	525
Db	407	AKEVKYKLEKCKTEKLWFCFDVIDIHTKATTKKEDIITKLFEFLFKPHVTGDVTGD-	464
QY	526	RYPFHLDGMQRVGNGKEAYKWELCHKBGHPASVWQRSLYNVNLGLKAPQMWTPKETGYTE	585
Db	465	-----TTVSEKEK-----SSGAKRK-----TPKTSPTA	490
QY	586	LVKSLERNWKLIIRDGLAMVDKAKGLFUPEDENIREK	622
Db	491	GSSSKBRKAKSOKKSEFATKVVKYSLAHSDDSEEEK	527

RESULT 14

h-caldesmon - chicken
N:Alternate names: caldesmon, smooth muscle, calmodulin- and actin-binding protein
C/Species: Gallus gallus (Chicken)
C/Date: 27-Feb-1990 #sequence revision 27-Jun-1994 #text change 22-Jun-1999
C/Accession: A33430; A32642; A32445; A41064; A60461; PC2003; PX0022
R:Hayaashi, K.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.
Biochem. Biophys. Res. Commun. 164, 503-511, 1989
A:Title: Primary structure and functional expression of h-caldesmon complementary DNA.
A:Reference number: A33430; MUID:90026426; PMID:2803315
A:Accession: A33430
A:Molecule type: mRNA
A:Residues: 1-771 <HA>
A:Cross-references: GB:M28417; NID:G211895; PIDN:AAA48810.1; PID:G211896
A:Experimental source: Gizzard
A:Note: part of this sequence was confirmed by protein sequencing
R:Bryan, J.; Inai, M.; Lee, R.; Moore, P.; Cook, R.G.; Lin, W.G.
J. Biol. Chem. 264, 13873-13879, 1989
A:Title: Cloning and expression of a smooth muscle caldesmon.
A:Reference number: A32642; MUID:89340480; PMID:2760048
A:Accession: A32642
A:Molecule type: mRNA
A:Residues: 1-318,334-771 <BRY>
A:Cross-references: GB:J04968; NID:G212656; PIDN:AAA49067.1; PID:G212657
A:Note: the authors translated the codon GAA for residue 743 as Leu
A:Note: this alternative splice form is a high molecular weight caldesmon (h-caldesmon)
R:Hayaashi, K.; Yamada, S.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.
Biochem. Biophys. Res. Commun. 161, 38-45, 1989
A:Title: 35KDa fragment of h-caldesmon conserves two consensus sequences of the tropomyo-
A:Reference number: A32445; MUID:89273666; PMID:2730665
A:Accession: A32445
A:Molecule type: mRNA
A:Residues: 466-771 <HA2>
A:Cross-references: GB:M26684; NID:G211897; PIDN:AAA48811.1; PID:G211898

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 10:56:23 ; Search time 18 Seconds
(without alignments)
2192.732 Million cell updates/sec

Title: US-09-903-216-2

Perfect score: 4022

Sequence: 1 MAQRKNAKSSGSSSSSGSGS.....IVDVHPELTPQQRSLPAI 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3965.5	98.6	757	1 ASPH_HUMAN	Q12797 homo sapien
2	3172	78.9	754	1 ASPH_BOVIN	Q28056 bos taurus
3	203	5.0	700	1 TRDN_CANFA	P82179 canis fami
4	194	4.8	4835	1 MDNI_GIALA	O85t1 giardia lam
5	193	4.8	493	1 ECX1_YETMA	O8p8t8 methanosarc
6	189.5	4.7	1616	1 P200_MTCGE	Q49429 mycoplasma
7	187	4.6	705	1 TRDN_RABIT	Q28820 oryctolagus
8	181	4.5	728	1 CALD_CHICK	P12957 gallus gall
9	178.5	4.4	728	1 TRDN_HUMAN	Q13061 homo sapien
10	176.5	4.4	845	1 NFM_RAT	P12839 rattus norv
11	175	4.4	2004	1 MY53_HUMAN	Q92794 homo sapien
12	172.5	4.3	1240	1 YNJI_YEAST	P53935 saccharomyc
13	172	4.3	465	1 YHU6_YEAST	P38845 saccharomyc
14	172	4.3	630	1 YCF2_OENVI	P31569 oenothera v
15	169.5	4.2	592	1 LAM2_MOUSE	P21619 mus musculu
16	169.5	4.2	1233	1 YFT6_YEAST	P43597 saccharomyc
17	165	4.1	650	1 NUC1_YENLA	P20397 xenopus lae
18	164	4.1	1357	1 SPOF_SCHPO	Q10411 schizosacch
19	163	4.1	411	1 FK93_YEAST	P38911 saccharomyc
20	162.5	4.0	793	1 CALD_HUMAN	Q05682 homo sapien
21	161.5	4.0	848	1 NFM_MOUSE	P08553 mus musculu
22	161.5	4.0	1220	1 IF2E_HUMAN	O60841 homo sapien
23	161.5	4.0	1395	1 SP41_YEAST	P38904 saccharomyc
24	161	4.0	2663	1 CENE_HUMAN	Q02224 homo sapien
25	161	4.0	4910	1 MDN1_YEAST	Q12019 saccharomyc
26	160.5	4.0	721	1 YCF2_OENPI	P31568 oenothera p
27	159	4.0	3924	1 ANK2_HUMAN	Q01484 homo sapien
28	158.5	3.9	837	1 RA50_METTH	Q26640 methanobact
29	158.5	3.9	1102	1 YG49_SCHPO	O60184 schizosacch
30	158	3.9	795	1 CDL1_HUMAN	P21127 homo sapien
31	157.5	3.9	780	1 CDL2_HUMAN	Q9uq88 homo sapien
32	157.5	3.9	1357	1 KTN1_HUMAN	Q86up2 homo sapien
33	156.5	3.9	506	1 NPL3_HUMAN	Q99457 homo sapien

34	156.5	3.9	1658	1 YM67_YEAST	Q03661 saccharomyc
35	154.5	3.8	1828	1 MAP2_MOUSE	P20357 mus musculu
36	154	3.8	1549	1 TRHY_SHEEP	P22793 ovis aries
37	153.5	3.8	1898	1 TRHY_HUMAN	Q07283 homo sapien
38	152	3.8	434	1 YKI2_YEAST	P36080 saccharomyc
39	152	3.8	671	1 CHS5_YEAST	Q12114 saccharomyc
40	152	3.8	725	1 HS9B_BRARE	O57521 brachydanio
41	152	3.8	728	1 HS9A_CHICK	P11501 gallus gall
42	152	3.8	1001	1 RPGR_MOUSE	O9r0x5 mus musculu
43	152	3.8	1233	1 SMIA_BOVIN	O97593 bos taurus
44	152	3.8	1233	1 SMIA_HUMAN	Q14683 homo sapien
45	152	3.8	1282	1 BMS1_HUMAN	Q14692 homo sapien

ALIGNMENTS

RESULT 1
ASPH_HUMAN
ID ASPH_HUMAN STANDARD; PRT; 757 AA.
AC Q12797;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Asparyl/asparaginyl beta-hydroxylase (EC 1.14.11.16) (Aspartate beta-hydroxylase) (ASP beta-hydroxylase) (Peptide-aspartate beta-dioxygenase).
DE ASPH.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
OX [1]
RN [1]
RP SEQUENCE FROM N.A. MEDLINE=95121937; PubMed=7821814;
RX Koriath F., Gieffers C., Frey J.;
RT "Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase".
RL Gene 150:395-399(1994).
CC -!- FUNCTION: Specifically hydroxylates an Asp or Asn residue in certain epidermal growth factor-like (EGF) domains of a number of proteins.
CC -!- CATALYTIC ACTIVITY: Peptide L-aspartate + 2-oxoglutarate + O(2) = peptide 3-hydroxy-L-aspartate + succinate + CO(2).
CC -!- COFACTOR: Iron.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic reticulum.
CC -!- TISSUE SPECIFICITY: Detected in all tissues tested.
CC -!- PTM: MIGHT BE PROCESSED TO THE 56 kDa (AA 274-757) OR 52 kDa (AA 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY SIMILARITY).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U03109; AAA82108.1; -.
PIR; I38423; I38423.
Genew; HGNC:757; ASPH.
MIM; 600582; -.
GO; GO:0005789; C:endoplasmic reticulum membrane; TAS.
GO; GO:0005509; F:calcium ion binding; TAS.
GO; GO:0005489; F:electron transporter activity; TAS.
GO; GO:0004597; F:peptide-aspartate beta-dioxygenase activity; TAS.
GO; GO:0008307; F:peptide-aspartate constituent of muscle; TAS.
GO; GO:0006936; P:muscle contraction; TAS.
InterPro; IPR007943; Asp-B-hydro N.
InterPro; IPR007803; Asp_Arg_Hydrox.

DR InterPro; IPR008940; Prenyl_trans.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF05279; Asp-B-Hydro N; 1.
 DR Pfam; PF05118; Asp Arg_Hydrox; 1.
 KW Oxidoreductase; Dioxxygenase; Iron; Transmembrane; Signal-anchor;
 KW Endoplasmic reticulum.
 FT DOMAIN 1 54 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 55 75 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 76 757 LUMENAL (POTENTIAL).
 FT DOMAIN 13 20 POLY-SER.
 FT DOMAIN 323 332 POLY-LYS.
 FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 705 705 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 757 AA; 85498 MW; 1A79313A4934C430 CRC64;
 Query Match 98.6%; Score 3966.5; DB 1; Length 757;
 Best Local Similarity 99.2%; Pred. No. 3.4e-193;
 Matches 752; Conservative 2; Mismatches 3; Indels 1; Gaps 1;
 QY 1 MAQRKNAKSGNSGSGSGSTAGSSSPGARETKHGKHGKRGKGLSGTSTFTFMV 60
 DB 1 MAQRKNAKSGNSGSGSGSTAGSSSPGARETKHGKHGKRGKGLSGTSTFTFMV 60
 QY 61 IALLGVNTSVAVVWFDLVYEEVLKGIYDADGDGDFVDDAKVLLGLKERSTSPAVP 120
 DB 61 IALLGVNTSVAVVWFDLVYEEVLKGIYDADGDGDFVDDAKVLLGLKERSTSPAVP 120
 QY 121 PEAAPHTPEEQVPEAEFQNTIDEAKEQIQSLHEMVAHVEGDLQOEGPTGEPQ 180
 DB 121 PEAAPHTPEEQVPEAEFQNTIDEAKEQIQSLHEMVAHVEGDLQOEGPTGEPQ 180
 QY 181 QEDDEFLMATVDVDRFETLEPEVSHETSHYVEETVSDQCDQNMSEENPDSS 240
 DB 181 QEDDEFLMATVDVDRFETLEPEVSHETSHYVEETVSDQCDQNMSEENPDSS 240
 QY 241 PVVEDERLHDDTVTYQVVEQAVPEPLENEGIEITEVTAPDNEVDQSIVVEVSI 300
 DB 241 PVVEDERLHDDTVTYQVVEQAVPEPLENEGIEITEVTAPDNEVDQSIVVEVSI 300
 QY 301 FVVEEQEVPETNRTKDDPEQAKVKKKPKLNLKPKDKTKIAELDAEKLKRGKIEEA 360
 DB 301 FVVEEQEVPETNRTKDDPEQAKVKKKPKLNLKPKDKTKIAELDAEKLKRGKIEEA 360
 QY 361 VNAFELVRVYPOSPRARYKACEDDLAEKRSNEVIRGAIETVQVAVSLPDVADLLK 420
 DB 361 VNAFELVRVYPOSPRARYKACEDDLAEKRSNEVIRGAIETVQVAVSLPDVADLLK 420
 QY 421 LSLKRRSDRQQLGHMRGSLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKVVYEV 480
 DB 421 LSLKRRSDRQQLGHMRGSLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKVVYEV 480
 QY 481 LSVTPNDGFAKHYGFLTKAQNKAESIPYLKEGIESGDPDGTDDGRFYFHLGDAMQVGN 540
 DB 481 LSVTPNDGFAKHYGFLTKAQNKAESIPYLKEGIESGDPDGTDDGRFYFHLGDAMQVGN 540
 QY 541 KEAYKVELGKRGHPRASVQWSLYNVNGLKAPQWTPKETGYTELKSLERNWKLIRDE 600
 DB 541 KEAYKVELGKRGHPRASVQWSLYNVNGLKAPQWTPKETGYTELKSLERNWKLIRDE 600
 QY 601 GLAVMDKAKGLFLPEDENLRKGDWSQFTLWQQGRNENACKGAPKCTCTLLEKFPETTC 660
 DB 601 GLAVMDKAKGLFLPEDENLRKGDWSQFTLWQQGRNENACKGAPKCTCTLLEKFPETTC 660
 QY 661 RRGQIKYSIMHPGTHVPHGTNCRMLHGLVPIKGECKIRCANETRWEEGKVLIFD 720
 DB 661 RRGQIKYSIMHPGTHVPHGTNCRMLHGLVPIKGECKIRCANETRWEEGKVLIFD 720
 QY 721 DSFEHEVQDASSRLLIFVDVWHPELTPOQRSLPAI 758
 DB 720 DSFEHEVQDASSRLLIFVDVWHPELTPOQRSLPAI 757

RESULT 2
 ASPH BOVIN STANDARD; PRT; 754 AA.
 ID ASPH BOVIN Q28056;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Aspartyl/asparaginyl beta-hydroxylase (EC 1.14.11.16) (Aspartate beta-hydroxylase) (ASP beta-hydroxylase) (Peptide-aspartate beta-hydroxylase).
 DE DE dioxxygenase).
 GN ASPH.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Brain, and Liver;
 RX MEDLINE=92332546; PubMed=1378441;
 RA Jia S., Vandusen W.J., Diehl R.E., Kohl N.E., Dixon R.A.F., Elliston K.O., Stern A.M., Friedman P.A.;
 RT "cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.";
 RL J. Biol. Chem. 267:14322-14327(1992).
 RN [2]
 RP SEQUENCE OF 289-385 AND 615-641.
 RC TISSUE=Liver;
 RX MEDLINE=91310699; PubMed=1856229;
 RA Wang Q., Vandusen W.J., Petroski C.J., Garsky V.M., Stern A.M., Friedman P.A.;
 RT "Bovine liver aspartyl beta-hydroxylase. Purification and characterization.";
 RL J. Biol. Chem. 266:14004-14010(1991).
 CC -!- FUNCTION: Specifically hydroxylates an Asp or Asn residue in certain epidermal growth factor-like (EGF) domains of a number of proteins.
 CC -!- CATALYTIC ACTIVITY: Peptide L-aspartate + 2-oxoglutarate + O(2) = Peptide 3-hydroxy-L-aspartate + succinate + CO(2).
 CC -!- COFACTOR: Iron.
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic reticulum.
 CC -!- PTM: Might be processed to the 56 kDa (AA 289-754) or 52 kDa (AA 311-754) forms in the lumen of the endoplasmic reticulum.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M91213; AAA03563.1; -.
 DR PIR; A42969; BABOH.
 DR InterPro; IPR007943; Asp-B-Hydro N.
 DR InterPro; IPR007803; Asp Arg_Hydrox.
 DR InterPro; IPR008941; TPR-like.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF05279; Asp-B-Hydro N; 1.
 DR Pfam; PF05118; Asp Arg_Hydrox; 1.
 KW Oxidoreductase; Dioxxygenase; Iron; Transmembrane; Signal-anchor;
 KW Endoplasmic reticulum.
 FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 58 78 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 79 754 LUMENAL (POTENTIAL).
 FT DOMAIN 9 12 POLY-GLY.
 FT DOMAIN 14 21 POLY-SER.
 FT DOMAIN 318 328 POLY-LYS.
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 754 AA; 84998 MW; 369593A1F0B558C8 CRC64;

Query Match
Best Local Similarity 78.9%; Score 3172; DB 1; Length 754;
Matches 612; Conservative 45; Mismatches 77; Indels 44; Gaps 6;

QY 1 MAOKKAKSSG--NSSSSGSGSGSTAGSSPGARRETHKGHGKNGKGLSGTSFPTW 57
Db 1 MAPKNAKGGGSSSSSSGSPCTSGSSSPGARRETHKGHGKNGKGLSGTSFPTW 60
QY 58 FMVIALGLWTSVAVWFDLVYEVVL-----GKLGIVYDADGDGDFVDD 102
Db 61 FMVIALGLWTSVAVWFDLVYEVVLAKAFRYNLSEVLQKLGIVYDADGDGDFVDD 120
QY 103 AKVLLGLKERSTSPAPVPEAEPTPEBQVPEAEAPONIEDEAKGQIQSLHEMHVAAE 162
Db 121 AKVLLGLKERSTSPAPVPEAEPTPEBQVPEAEAPONIEDEAKGQIQSLHEMHVAAE 162
QY 121 AKVLLGLKERSTSPAPVPEAEPTPEBQVPEAEAPONIEDEAKGQIQSLHEMHVAAE 162
Db 121 AKVLLGLKERSTSPAPVPEAEPTPEBQVPEAEAPONIEDEAKGQIQSLHEMHVAAE 162
QY 163 HVEGEDIQEDGPTGEQOEDDFLMATDVRDFTLEPEVSHVETSHVETVSQDC 222
Db 180 --PGENUPQEPGPAELQPDHVFVGSADDRPEMGTVGAVHEETDSYHIEETASPAY 237
QY 223 NQDEEMWSEQENPDSEPPVVED--ERLHHDTDVTVQVVEEQAVYEPLENEGIEITEVT 280
Db 238 SQMEDMAYEQENPDSEPPVVEDAERTYQETDDVTVQVVEEQAVYEPLENEGIEITEVT 280
QY 281 APPENPDVEDSQVIVEEVSIFPVVEEQVPEPTNRKTDPPQAKVKKKPKLLINKDKT 340
Db 281 ----DHAVDNTTILEEPHMPAEQEQVPEPTNRKTDPPQAKVKKKPKLLINKDKT 336
QY 341 IKAELDAEKLKRGKTEAEVNAFKELVRKYPQSPRARYGKACEDDLAEKRRSNEVLRG 400
Db 337 IKAELDAEKLKRGKTEAEVNAFKELVRKYPQSPRARYGKACEDDLAEKRRSNEVLRG 396
QY 401 AIETYQEVASLPDVPADLLKLSKRRSDRQFLGHMRGSLTLQELVOLFPNDTSLKNDL 460
Db 397 AIETYQEVASLPDVPADLLKLSKRRSDRQFLGHMRGSLTLQELVOLFPNDTSLKNDL 456
QY 461 GVGILLGDNDNAKAYVEEVLVTPNDGFAKHGFIKAQNKIAESTPYLKEGIESGDP 520
Db 457 GVGILLGDNDNAKAYVEEVLVTPNDGFAKHGFIKAQNKIAESTPYLKEGIESGDP 516
QY 521 GTDGRGFYHLGDAMQRYGKAEYKWLHGRGHFASVWQSLYVNLGLAKAPWWTPE 580
Db 517 GTDGRGFYHLGDAMQRYGKAEYKWLHGRGHFASVWQSLYVNLGLAKAPWWTPE 576
QY 581 TGYTELKSLERNWKLIRDEGLAVMDKAGLFLPEDENLREKGDWSQFTLMQOGRNENA 640
Db 577 TGYTELKSLERNWKLIRDEGLAVMDKAGLFLPEDENLREKGDWSQFTLMQOGRNENA 636
QY 641 CKGAPKCTILLEKFPETTCGRGQIKYSTHMPGTHVHTGTCNCLBMHLGLVLPKGC 700
Db 637 CKGAPKCTILLEKFPETTCGRGQIKYSTHMPGTHVHTGTCNCLBMHLGLVLPKGC 696
QY 701 KIRCANERTWEEGKVLFDSDFEHEVWQDASFRLLIFVDVWHPHLPQQRSLPAI 758
Db 697 KIRCANERTWEEGKVLFDSDFEHEVWQDASFRLLIFVDVWHPHLPQQRSLPAI 754

RESULT 3
ID TRDN CANFA STANDARD; PRT; 700 AA.
AC P82179;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Triadin.
GN TRDN.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
```

Qy		136	V E A --- E P O N I E D A K Q I O S L H M V H A E H V E G E D I Q O E D G P T G B F Q C Q O E D F I M A I D V	192
Dy		160	A K V A H R E K E V X E K E S E --- K K A T H K E I E K E K P E T K T M A K E R K A T E E K I K E V	214
Qy		193	- D R F E T L E P E V S --- H E R T E H --- S Y H V E E T V S O D C	222
Dy		215	K G G Q E K V K P T A A K V E Q V T P P K A K E K G E T A A V A K H E Q Q A F C R Y M I D M F V H G D L	274
Qy		223	N O D M E E N W --- S Q E N P D S E P V V E D R L H H D T D D V T Y Q V Y E	261
Dy		275	R P G Q S P A L P P L P T V Q A S R P T P A S P T L E G K E E E K K A E K K V T S E T K K K E K E D V K K G S D K	334
Qy		262	E Q A V --- Y P L E N E G - I E T E V T A P E D N P V E S Q V I V E V S F P P V E E Q O E ---	308
Dy		335	D T A I D V B K E G K A P E T K Q G T I K V V A Q A A K D E K K D S K T K T P V E S H P K G K O E K K E K	394
Qy		309	- V P P E T N R K T D D --- P E O K --- A K V K K K K P - K L I N K F D K T I K A E L D A A E	349
Dy		395	Y V E P A K S K K E H S A P S E Q V K A K T E R A K E E T S A S T K K A V P G K K E K T T V Q E B I ---	450
Qy		350	K L R K R G K I --- E E A V N A F K E L V R X Y P O S P R A R Y G K --- A Q C E D D	387
Dy		451	R K E S G K T S T A S K D K P E I K X D K M P K A D K V K P P O S Q V K K E K S E S Q V K K A K P E Q D	510
Qy		388	L A E K R S --- N E V U R G A I E --- T Y Q E V A S L P D V P A D L L K L S	422
Dy		511	I A K P E K T V S H G K P E E K V V K Q V K A T E A A I K T V K P K A K A B H O K E S - P T I K T D K P R P T	569
Qy		423	L K --- R S D R Q O F I G H M R G S L L T I Q R L V O L P N D T S L K N D L G V G Y L	465
Dy		570	S K E P T V E T S G K K I E K S E K S E K A B M K H L K E E K V S T R K E S L Q S H N V T K A E K P A R V S R E	629
Qy		466	L I G N D N A K V Y E E V L S V --- T P N D G F A K V H --- Y G F I L K A Q N K I A E S I P Y L K	512
Dy		630	D L E D V S A K K A E A E A D V S S T K R O K S P I S F F Q C V Y L D G Y N G Y G F P P V --- T P A Y R	682
Qy		513	E G I E S G D P G T	522
Dy		683	P G E S S G Q P S S	692

```

RESULT 4
MDN1_GIALA MDN1_GIALA STANDARD; PRT; 4835 AA.
AC Q875T1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Midasin (MIDAS-containing protein).
GN MDN1.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20389616; PubMed=10930750;
RA McArthur A.G., Morrison H.G., Nixon J.E., Passamaneck N.Q., Kim U.,
RA Hinkle G., Crocker M.K., Holder M.E., Farr R., Reich C.I., Olsen G.E.,
RA Aley S.B., Adam R.D., Gillin F.D., Sogin M.L.;
RT "The Giardia genome project database.";
RL FEMS Microbiol. Lett. 189:271-273(2000).
RN [2]
RP IDENTIFICATION, GENE NAME, AND SIMILARITY WITH OTHER FAMILY MEMBERS.
RX PubMed=12102729;
RA Garbarino J.E., Gibbons I.R.;
RT "Expression and genomic analysis of midasin, a novel and highly
RT conserved AAA protein distantly related to dynein.";
RL BMC Genomics 3:18-18(2002).
CC -!- FUNCTION: May function as a nuclear chaperone and be involved in
CC the assembly/disassembly of macromolecular complexes in the
CC nucleus.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 1 VWFA domain.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF494287; AAM12656.1; --
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003599; AAA_ATPase_cent.
CC InterPro; IPR002035; VWF_A.
CC Pfam; PF00004; AAA; 2.
CC Pfam; PF00004; AAA; 2.
CC SMART; SM00382; AAA; 5.
CC SMART; SM00327; VWA; 1.
CC PROSITE; PS50234; VWF; 1.
CC DR Chapterone; ATP-binding; Repeat; Nuclear protein.
CC KW Chapterone; ATP-binding; Repeat; Nuclear protein.
CC FT NP BIND 356 363 ATP (POTENTIAL) .
CC FT NP BIND 814 821 ATP (POTENTIAL) .
CC FT NP BIND 1127 1134 ATP (POTENTIAL) .
CC FT NP BIND 1513 1520 ATP (POTENTIAL) .
CC FT NP BIND 1839 1846 ATP (POTENTIAL) .
CC FT NP BIND 3277 3284 ATP (POTENTIAL) .
CC FT NP BIND 110 113 POLY-LEU.
CC FT DOMAIN 3004 3007 POLY-LEU.
CC FT DOMAIN 4156 4161 POLY-ASN.
CC FT DOMAIN 4629 4818 VWFA.
CC FT SEQUENCE 4835 AA; 539726 MW; 3A9E12417DB04A50 CRC64;
CC -----
CC Query Match 4.8%; Score 194; DB 1; Length 4835;
CC Best Local Similarity 23.4%; Pred. No. 0.084;
CC Matches 119; Conservative 74; Mismatches 203; Indels 120; Gaps 25;
CC -----
CC QY 80 YEEVLGKLGKLYD--ADGGDFDVEDDAKVLGLKERSTSEPAVPEEAEPTPEEQVPE 137
CC DB 4188 HEEQADATGSTDQAQEDDYNDLDD-KNLSG-----QSDLSVPEADGEDETVNEE---LE 4238
CC QY 138 AEPONIDEAKEQIQSLLEHVMHAHVHGEDLQOEGDPTGPEQOEDDEFMATDVEDRFE 197
CC DB 4239 EEQQQSGDLSPND-----QDACAIEEDDDRDLPSSD---ENAEHEHDEAPVDIDDN-E 4288
CC QY 198 TLEPVSHHEETSHYHVEETVSDCNODEMWMSEQE-----NPDSEPVVEDER- 247
CC DB 4289 ASDEQSTYNDRDDDAINIQAQQATNDEEMQKDTEDQENITDSNPDANEVGTNDQKQ 4348
CC QY 248 LHHDTDVTVQVYBEQAVYPLEN---EGIEITVTPAPPNDPNVEDSQVI-VREVSTFPV 303
CC DB 4349 THEDNQFRQENIBDWEAEASTENSQGEAESAADLKEGNDPMSLEEFQIRWIKERLNIHDR 4408
CC QY 304 E-----EQQVPEPTNRKTDDEQKA-----KVKKKKPKLLNKFDTIKAEIDAAE 349
CC DB 4409 ESEKDEAAEPQMPQLQNKTVFDDSKSGRDGALGLTESKRNLTNQ-----EFQNP 4461
CC QY 350 KLRKRGKIEEAVNAFKELVRKYPOSRYAKCAQCEDDLAP-----KRR 393
CC DB 4462 EER---NVEH--NSSCETSSQSHDRPAPAEHLNPEISDEGEESSTASKQEQAVLSHMRES 4516
CC QY 394 SNEVIRGAITYQYQVA-SLPD-----VPADI-----LKLSLKRSRDEQQLGHMRG 438
CC DB 4517 SKDLINPEGEVYQELAVSLASEETKRAPEDVAAASARGNHLDDLIIKOTSAAAP----- 4570
CC QY 439 SLILTLQRLVOLFPNDTS-LKNDLGVYLL-----IGNDNNAKKVYE 478
CC DB 4571 SLAELRLII-LEPTVTSDLKGDFTGKKLNLRILPIPTASEFQDKIWLRRTKFSKRVYQ 4629
CC QY 479 EVLSVPTNDGFAKVHYGFILKAQNKIAESIPYLKEG 514
CC DB 4630 VLLAVDDSSMAPI-AKVALQAILTLFNACKFLEVG 4664
CC -----
CC RESULT 5
CC ECX1 METWA

ID	ECX1 METMA	STANDARD;	PRT;	493 AA.
AC	Q8PTT8;			
DT	10-OCT-2003 (Rel. 42, Created)			
DT	10-OCT-2003 (Rel. 42, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Probable exosome complex exonuclease 1 (EC 3.1.13.-).			
GN	MX2623.			
OS	Methanosarcina maezi (Methanosarcina frisia).			
OC	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;			
OC	Methanosarcinaceae; Methanosarcina.			
OX	NCBI_TaxID=2209;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Goel / GoI / ATCC BAA-199 / DSM 3647 / OCM 88;			
RX	MEDLINE=22120827; PubMed=12125824;			
RA	Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,			
RA	Martinez-Arias R., Henne A., Wierse A., Baeumer S., Jacob C.,			
RA	Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,			
RA	Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,			
RA	Fritz H.-J., Gottschalk G.;			
RT	"The genome of Methanosarcina maezi: evidence for lateral gene			
RT	transfer between Bacteria and Archaea.";			
RL	J. Mol. Microbiol. Biotechnol. 4:453-461(2002).			
CC	- FUNCTION: Probably involved in the 3'->5' degradation of a variety			
CC	of RNA species (Potential).			
CC	- SUBUNIT: Component of the archaeal exosome multienzyme			
CC	ribonuclease complex (Potential).			
CC	- SUBCELLULAR LOCATION: Cytoplasmic (Potential).			
CC	- SIMILARITY: Belongs to the RNase PH family.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC				
DR	EMBL; AB013507; AAM32319.1; -			
DR	HMAP; MF_00591; -; 1.			
DR	InterPro; IPR001247; 3 ExonNase.			
DR	Pfam; PF01138; RNase PH; 1.			
DR	Pfam; PF01725; RNase PH C; 1.			
KW	Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.			
FT	DOMAIN 1 254 PROBABLE EXOSOME COMPLEX EXONUCLEASE 1.			
FT	DOMAIN 255 493 UNKNOWN.			
FT	SEQUENCE 493 AA; 55248 MW; 979A757BEF8DC090 CRC64;			
Query Match				
Best Local Similarity	4.8%; Score 193; DB 1; Length 493;			
Matches	97; Conservative 45; Mismatches 129; Indels 98; Gaps 17;			
Qy	61 IALLGWTSVAVVFDVYEVVLGKLGIDYADGGDFV-----DDAKVLL----- 107			
Db	153 IPMKGLITSCA---FGKVDGKIVLNDKEDNYGEADFPVAMTQDGEITLIQMDGNLTPD 209			
Qy	108 -----GLKRSSTSEPAV-----PPEAEPHTPE-----EQVPEAPEQ 141			
Db	210 EIKQGLVKKCKEILEIQQAVLRKFTPEVSEETAPEKGAKEVLEPSPAIVE 269			
Qy	142 NTEDEAKEIQSLLHMHVHAEVGEEDLQOEDGPTGEPOQED--DEFLMAT--DVDDRPE 197			
Db	270 ETPPEAEP-----EVEISEEVEAELASVLPDPEDELEIEEIEELEESEEDLETEE 323			
Qy	198 TLEPEVSHETSHYVETVQDCNQDMEMWSEQNPDSSEPVVEDE----- 246			
Db	324 EFEEAELEAEPPEDELEEDLGELEEEEEELEEEFEAELEEELEAEELEAEELECAP 383			
Qy	247 -----RLHDTDDVYQVYE-----EQAVVEPLENGIEITETVAPPDNPVED 290			
Db	384 ELKEFDEIEARL--EKEDASIEAEERIEPEAEATEEGLEEEA-EIEETAASEEN-IEA 439			
Qy	291 SQVIVEVSIFPVVEQQEVPETNRKTDDEPEQAKVKKKKPKLLNFKDKTKIAELDAEK 350			

Db	440 EAAEAEAE--FEVAAEISTEAEAEAEPEPE---KSEGPW-----KVKDPSAG-- 486			
Qy	351 LKRGKIEE 359			
Db	487 --TRGEKDE 493			
RESULT 6				
P200_MYCGE				
ID	P200_MYCGE	STANDARD;	PRT;	1616 AA.
AC	Q49429; Q49259; Q49298; Q49352; Q49353;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Protein P200.			
GN	MG386.			
OS	Mycoplasma genitalium.			
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.			
OX	NCBI_TaxID=2097;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 33530 / G-37;			
RX	MEDLINE=96026346; PubMed=7569993;			
RA	Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,			
RA	Fleischmann J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,			
RA	Fritchman R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,			
RA	Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,			
RA	Tomb J.-F., Dougherty B.A., Bost K.F., Hu P.-C., Lucier T.S.,			
RA	Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;			
RT	"The minimal gene complement of Mycoplasma genitalium.";			
RL	Science 270:397-403(1995).			
EN	[2]			
RP	SEQUENCE OF 256-427; 432-543 AND 1083-1140 FROM N.A.			
RX	STRAIN=ATCC 33530 / G-37;			
RC	MEDLINE=94075230; PubMed=8253680;			
RX	Peterson S.N., Hu P.-C., Bost K.F., Hutchison C.A. III;			
RA	"A survey of the Mycoplasma genitalium genome by using random			
RT	sequencing.";			
RL	J. Bacteriol. 175:7918-7930(1993).			
CC	- FUNCTION: COULD BE AN ACCESSORY STRUCTURAL COMPONENT IN			
CC	CYTADHERENCE (BY SIMILARITY).			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC				
DR	EMBL; U39720; AAC71613.1; -			
DR	EMBL; U02245; AAC03400.1; -			
DR	EMBL; U02245; -; NOT ANNOTATED_CDS.			
DR	EMBL; U02175; AAD12458.1; -			
DR	EMBL; U02126; AAD12402.1; -			
DR	PIR; G64242; G64242.			
DR	TIGR; MG386; -			
KW	Cytadherence; Structural protein; Repeat; Complete proteome.			
FT	DOMAIN 1205 1389 2 X 32 AA REPEAT.			
FT	REPEAT 1205 1389 1-1.			
FT	REPEAT 1358 1389 1-2.			
FT	DOMAIN 891 1389 2 X 26 AA REPEAT.			
FT	REPEAT 1161 1186 2-1.			
FT	REPEAT 1310 1339 2-2.			
FT	CONFLICT 256 256 P -> S (IN REF. 2).			
FT	CONFLICT 304 304 S -> F (IN REF. 2).			
FT	SEQUENCE 1616 AA; 185678 MW; 6AF76A13AC49E4FF CRC64;			
Query Match				
Best Local Similarity	4.7%; Score 189.5; DB 1; Length 1616;			
Matches	142; Conservative 78; Mismatches 205; Indels 213; Gaps 35;			


```
QY 75 FDLVDYEEVLGK-----LGIYD-----ADGSDGDFDVKVL----- 106
Db 1091 FDTVRHEAVDFKQOTQEGLEBPQVSSBAEVVDQTTTIVGPEAVDFDQPEKTEVKEFD 1150
QY 107 -LGLKERSTSEPAVPEAPHTPEEQVPEABPNQIEDAKBQIQSLHHEVHAHVYE 165
Db 1151 DVNQKQVISEPQVQEQGAEVPEPSAEAKFDSQVSDQSPQVPEVLEEVQTPQEPV 1210
QY 166 GEDLQOEDGPTGEPQOEDDEFLMATDVRDFETLE-PEVSHETESHYHVEEIVSOD-CN 223
Db 1211 SQPEATFD--TVQEQTPQE-----AKFDSQVETVEQPEFSSEPTQO--HVESASFDEN 1262
QY 224 QDMEMMSQENP--DS-----SEPVEDERLHHDTDVITYQV-----YEQAVYEPL 269
Db 1263 YDEDENYDFDQSYSDSLQPSQVVDPEPNVDPEPNVEIESKSEPOFQVQEQP- 1321
QY 270 ENGIEITVTAPE--DNPVE-----DSQ--VIVEVSIFPVEEQOEVPPFTNRTYDDPEQK 323
Db 1322 --GEAVFSPSAEAKFDSQVSDQSPQVPEVLEEVQTPQEPVSDQSPQVPEVLEEVQTPQEP 1378
QY 324 AKVKKKPKLLNKFDTKIAELDAEAKLRKRGKIEAVNAFVKELVRKYPOSPEARYKQ 383
Db 1379 PQEA-----KFD-----SP----- 1387
QY 384 CEDDLAEKRRSNEVLRAIETQOE--VASLPDVPADLLKLSKRRSDROQFLGHRGSL 441
Db 1388 -----VETIQEQVSSPEV--VVQPNFEERKPE-----TVL 1417
QY 442 TLQRLVOLFP--NDTSLKNDLGVYLLIGDNNAKVYEVSVTPNDQFAKHVGFIL 498
Db 1418 EEPQADIEPEASEESLDWELLVG-----NNSYGHYEP-----DG-EWVWAGFFG 1462
QY 499 KAO--NKIAPSIPLYKIEGSDPGDGRFYFHL-QDAMQ-RVGNKEAYKVELGHRG 554
Db 1463 DDQWKNDA-TVXWRE-----RDYPLIGDEVYGRYNNKGEWINGFYDES 1509
QY 555 HFASV---WQSLYNNVGLKAQPMWTP-----KETGYTELKSLRNWKLIRDEGLAVMDK 607
Db 1510 DWLVDVDEQWKNRPRIN--EAPKFEKLIQNEEYGYE--DNEWNW----- 1551
QY 608 AKGLFLPEDENLRKGDWSOFTLWQGRNRNACKGAP 645
Db 1552 -----YDGEFDSQGNLWVQSEETENLADITKDIP 1582

RESULT 7
TRDN_RABIT STANDARD; PRT; 705 AA.
AC Q28820; Q28636; Q28637; Q28643;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DB Triadin.
GN TRDN.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).
RC STRAIN=New Zealand white; TISSUE=Skeletal muscle;
RA MEDLINE=93286104; PubMed=7685347;
RX Knudson C.M., Stang K.K., Moonaw C.R., Slaughter C.A., Campbell K.P.;
RT "Primary structure and topological analysis of a skeletal muscle-
specific junctional sarcoplasmic reticulum glycoprotein (triadin).";
RL J. Biol. Chem. 268:12646-12654 (1993).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).
RC TISSUE=Skeletal muscle;
RX MEDLINE=94298946; PubMed=8026576;
RA Peng M., Fan H., Kirley T.B., Caswell A.H., Schwartz A.;
RT "Structural diversity of triadin in skeletal muscle and evidence of
its existence in heart.";
```

```
FEBS Lett. 348:17-20 (1994).
[3]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS CARDIAC).
RC TISSUE=Heart muscle;
RX MEDLINE=96132942; PubMed=8550602;
RA Guo W., Jorgensen A.O., Jones L.R., Campbell K.P.;
RT "Biochemical characterization and molecular cloning of cardiac
triadin.";
RL J. Biol. Chem. 271:458-465 (1996).
[4]
RP CARBOHYDRATE-LINKAGE SITES, AND INTERCHAIN DISULFIDE BONDS.
RX MEDLINE=96066664; PubMed=7578102;
RA Fan H., Brandt N.R., Caswell A.H.;
RT "Disulfide bonds, N-glycosylation and transmembrane topology of
skeletal muscle triadin.";
RL Biochemistry 34:14902-14908 (1995).
CC -!- FUNCTION: May be involved in anchoring calsequestrin to the
junctional sarcoplasmic reticulum and allowing its functional
coupling with the ryanodine receptor.
CC -!- SUBUNIT: Homooligomer of variable subunit number; disulfide-
linked.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
reticulum.
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=6;
Comment=Additional isoforms seem to exist;
Name=Skeletal 1; Synonyms=ST1;
IsoId=Q28820-1; Sequence=Displayed;
Name=Cardiac 1; Synonyms=CTL1;
IsoId=Q28820-2; Sequence=VSP_004458, VSP_004460;
Name=Cardiac 2; Synonyms=CTL2;
IsoId=Q28820-3; Sequence=VSP_004459, VSP_004461;
Name=Cardiac 3; Synonyms=CTL3;
IsoId=Q28820-4; Sequence=VSP_004466;
Name=Skeletal 2; Synonyms=ST2;
IsoId=Q28820-5; Sequence=VSP_004462, VSP_004463, VSP_004464,
VSP_004465;
Name=Skeletal 3; Synonyms=ST3;
IsoId=Q28820-6; Sequence=VSP_004464, VSP_004465;
-!- TISSUE SPECIFICITY: Skeletal and cardiac muscle.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
EMBL; U31540; AAC48496.1; -
EMBL; L10065; AAA31488.1; -
EMBL; U31555; AAC48497.1; -
EMBL; U34201; AAC48498.1; -
DR PR; A45990; A45990.
DR Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
KW Alternative splicing.
FT INIT MET 0
FT DOMAIN 1 46 CYTOPLASMIC.
FT TRANSMEM 47 67 POTENTIAL.
FT DOMAIN 68 705 LUMENAL.
FT CARBOHYD 74 74 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 624 624 N-LINKED (GLCNAC...).
FT VARSPLIC 264 285 DOYAFRCYRMDIVFHGDLKPGQ -> GKGQSEBAAGCFKKT
LGHKQMQ (in isoform Cardiac 1).
FT FTId=VSP_004458.
FT DOYAFRCYRMDIVFHGDLKPGQSPAPPPSPPTQASRPPTA
FT LPT -> ECIFLSAATPQGPINRQQLNDIHCEFLTKYKGN
FT GQFAFLKGC (in isoform Cardiac 2).
FT FTId=VSP_004459.
FT Missing (in isoform Cardiac 1).
FT VARSPLIC 286 705 /FTId=VSP_004460.
FT VARSPLIC 308 705 Missing (in isoform Cardiac 2).
FT /FTId=VSP_004461.
```


regulation of actomyosin interactions in smooth muscle and nonmuscle cells (could act as a bridge between myosin and actin filaments). Stimulates actin binding of tropomyosin which increases the stabilization of actin filament structure. In muscle tissues, inhibits the actomyosin ATPase by binding to F-actin. This inhibition is attenuated by calcium-calmodulin and is potentiated by tropomyosin. Interacts with actin, myosin, two molecules of tropomyosin and with calmodulin. Also play an essential role during cellular mitosis and receptor coupling.

CELLULAR LOCATION: On thin filaments in smooth muscle and on stress fibers in fibroblasts (nonmuscle) (By similarity).

ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=3;

Name=Gizzard h-cad;

IsoId=P12957-1; Sequence=displayed;

Name=Brain l-cad;

IsoId=P12957-2; Sequence=VSP_004152, VSP_004153;

Name=Gizzard l-cad;

IsoId=P12957-3; Sequence=VSP_004153;

TISSUE SPECIFICITY: High-molecular-weight caldesmon (h-caldesmon) is predominantly expressed in smooth muscles, whereas low-molecular-weight caldesmon (l-caldesmon) is widely distributed in non-muscle tissues and cells. Not expressed in skeletal muscle or heart.

DOMAIN: The N-terminal part seems to be a myosin/calmodulin-binding domain, and the C-terminal a tropomyosin/actin/calmodulin-helical region in the muscle forms.

PTM: Phosphorylated in non-muscle cells. Phosphorylation by Cdc2 during mitosis causes caldesmon to dissociate from microfilaments. Phosphorylation reduces caldesmon binding to actin, myosin, and calmodulin as well as its inhibition of actomyosin ATPase activity. Phosphorylation also occurs in both quiescent and dividing smooth muscle cells with similar effects on the interaction with actin and calmodulin and on microfilaments reorganization (By similarity).

SIMILARITY: Belongs to the caldesmon family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; J04968; AAA49067.1; -
 EMBL; D17648; BAA04539.1; -
 EMBL; D17634; BAA04539.1; JOINED.
 EMBL; D17636; BAA04539.1; JOINED.
 EMBL; D17637; BAA04539.1; JOINED.
 EMBL; D17638; BAA04539.1; JOINED.
 EMBL; D17639; BAA04539.1; JOINED.
 EMBL; D17640; BAA04539.1; JOINED.
 EMBL; D17641; BAA04539.1; JOINED.
 EMBL; D17642; BAA04539.1; JOINED.
 EMBL; D17643; BAA04539.1; JOINED.
 EMBL; D17644; BAA04539.1; JOINED.
 EMBL; D17645; BAA04539.1; JOINED.
 EMBL; D17646; BAA04539.1; JOINED.
 EMBL; M28417; AAA48810.1; -
 EMBL; M60620; AAA48936.1; -
 EMBL; D17648; BAA04538.1; -
 EMBL; D17635; BAA04538.1; JOINED.
 EMBL; D17636; BAA04538.1; JOINED.
 EMBL; D17637; BAA04538.1; JOINED.
 EMBL; D17639; BAA04538.1; JOINED.
 EMBL; D17640; BAA04538.1; JOINED.
 EMBL; D17641; BAA04538.1; JOINED.
 EMBL; D17642; BAA04538.1; JOINED.
 EMBL; D17643; BAA04538.1; JOINED.
 EMBL; D17644; BAA04538.1; JOINED.
 EMBL; D17645; BAA04538.1; JOINED.

regulation of actomyosin interactions in smooth muscle and nonmuscle cells (could act as a bridge between myosin and actin filaments). Stimulates actin binding of tropomyosin which increases the stabilization of actin filament structure. In muscle tissues, inhibits the actomyosin ATPase by binding to F-actin. This inhibition is attenuated by calcium-calmodulin and is potentiated by tropomyosin. Interacts with actin, myosin, two molecules of tropomyosin and with calmodulin. Also play an essential role during cellular mitosis and receptor coupling.

CELLULAR LOCATION: On thin filaments in smooth muscle and on stress fibers in fibroblasts (nonmuscle) (By similarity).

ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=3;

Name=Gizzard h-cad;

IsoId=P12957-1; Sequence=displayed;

Name=Brain l-cad;

IsoId=P12957-2; Sequence=VSP_004152, VSP_004153;

Name=Gizzard l-cad;

IsoId=P12957-3; Sequence=VSP_004153;

TISSUE SPECIFICITY: High-molecular-weight caldesmon (h-caldesmon) is predominantly expressed in smooth muscles, whereas low-molecular-weight caldesmon (l-caldesmon) is widely distributed in non-muscle tissues and cells. Not expressed in skeletal muscle or heart.

DOMAIN: The N-terminal part seems to be a myosin/calmodulin-binding domain, and the C-terminal a tropomyosin/actin/calmodulin-helical region in the muscle forms.

PTM: Phosphorylated in non-muscle cells. Phosphorylation by Cdc2 during mitosis causes caldesmon to dissociate from microfilaments. Phosphorylation reduces caldesmon binding to actin, myosin, and calmodulin as well as its inhibition of actomyosin ATPase activity. Phosphorylation also occurs in both quiescent and dividing smooth muscle cells with similar effects on the interaction with actin and calmodulin and on microfilaments reorganization (By similarity).

SIMILARITY: Belongs to the caldesmon family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; D17646; BAA04538.1; JOINED.
 DR EMBL; D17647; BAA04538.1; JOINED.
 DR EMBL; D17648; BAA04540.1; -
 DR EMBL; D17649; BAA04540.1; -
 DR EMBL; D17634; BAA04540.1; JOINED.
 DR EMBL; D17636; BAA04540.1; JOINED.
 DR EMBL; D17637; BAA04540.1; JOINED.
 DR EMBL; D17638; BAA04540.1; JOINED.
 DR EMBL; D17639; BAA04540.1; JOINED.
 DR EMBL; D17640; BAA04540.1; JOINED.
 DR EMBL; D17641; BAA04540.1; JOINED.
 DR EMBL; D17642; BAA04540.1; JOINED.
 DR EMBL; D17643; BAA04540.1; JOINED.
 DR EMBL; D17644; BAA04540.1; JOINED.
 DR EMBL; D17645; BAA04540.1; JOINED.
 DR EMBL; D17646; BAA04540.1; JOINED.
 DR EMBL; D17647; BAA04540.1; JOINED.
 DR EMBL; M59762; AAA48649.1; -
 DR EMBL; D17552; BAA04490.1; -
 DR EMBL; M26684; AAA48811.1; -
 DR PIR; A33430; A33430.
 DR HSSP; P05412; 1FOS.
 DR InterPro; IPR006017; Caldesmon.
 DR InterPro; IPR006018; Caldesmon_LSP.
 DR Pfam; PF02029; Caldesmon; 1.
 DR PRINTS; PR01076; CALDESMON.
 KW Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;
 KW Repeat; Alternative splicing.
 FT DOMAIN 26 199 MYOSIN AND CALMODULIN-BINDING.
 FT REPEAT 251 390 10 X 13 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 251 265 1.
 FT REPEAT 266 278 2.
 FT REPEAT 279 291 3.
 FT REPEAT 294 306 4.
 FT REPEAT 309 321 5.
 FT REPEAT 324 336 6.
 FT REPEAT 337 349 7.
 FT REPEAT 350 362 8.
 FT REPEAT 363 375 9.
 FT REPEAT 378 390 10.
 FT DOMAIN 523 580 TROPOMYOSIN-BINDING (POTENTIAL).
 FT DOMAIN 622 632 TROPOMYOSIN-BINDING (POTENTIAL).
 Query Match 4.5%; Score 181; DB 1; Length 771;
 Best Local Similarity 22.8%; Pred. No. 0.039;
 Matches 150; Conservative 87; Mismatches 248; Indels 174; Gaps 32;

Qy 108 GLKERSTSE---PAVPPEE-----AEPHTEPEEQV-----PVEAE-----PQNI 143
 Db 178 GKKEEKDSEEEKPEVTEENQVDVAVEKSTDKEEVVETKTLVNAENDTNAMLEGQSI 237
 Qy 144 EDEAKEIQISLLHEMVHAHVEGEDLQ-QEDGPTGPEQEDDFLMTVDVDRFETLEPE 202
 Db 238 TDAADKEKEAEKEREKLEAEKEREKLEAEKEREKLEAEKEREKLEAEKEREKLEAEKEREK 297
 Qy 203 VSHETEHSYHVEETVSQDCNQDMENMSEQNPDSSEPV-VEDERLHHTDDVTVQVYE 261
 Db 298 RAAAEERERAKAEAE---RKAAEERERAKAEAEERAKAEAEERAKAEAEERAKAEAEER 344
 Qy 262 EQAVVEPLENEGTEITETVAPPEDNPVEDSQVVEEVSIFPVEEQVEPPETNR----- 315
 Db 345 ERAKAE-----ERKAEERERAKAEERERAKAEERERAKAEERERAKAEERERAKAE 395
 Qy 316 --RTDDPEQAKVKKKKPKLLNPKTKTKAELDA-----AEKL----- 351
 Db 396 KEKKKEKKAQAEKKAQANLLRQEDKEAKVEAKESLPEKLPQTSKQDQVKNKDKKEK 455
 Qy 352 -----RKCKTIE-AVNNAFKELVKYQSPARYKQACEDDLAEKRSNEVLK 399
 Db 456 APKEEMKSVMDRKRGVPEQKAQNGERELTTPKLSKTENAFGRSNLK-GAANAEEAGSKLK 514
 Qy 400 GAITYQVAVSLPDVPADLLKLSLRKRSRQOQFGLHMRGSLITLQRLVQLFPNDTSLKND 459
 Db 515 ---EKQERAA-----VELDEUK---KRREERKIL-----ESEEQKKK 546

EMBL; U47742; AAC50662.1; -.
 DR Genew; HGNC:13013; MYST3.
 DR MIN; 601408; -.
 DR GO; GO:0006323; P:DNA packaging; TAS.
 DR InterPro; IPR005818; Histone.H1/H5.
 DR InterPro; IPR002717; MOZ_SAS.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF01853; MOZ_SAS; 1.
 DR Pfam; PF00628; PHD; 2.
 DR SMART; SM00526; H15; 1.
 DR SMART; SM00249; PHD; 2.
 DR PROSITE; PS01359; ZF_PHD_1; 1.
 DR PROSITE; PS00016; ZF_PHD_2; 2.
 DR Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
 KW Nuclear protein.
 FT ZN_FING 206 265 PHD-TYPE 1.
 FT ZN_FING 259 313 PHD-TYPE 2.
 FT DOMAIN 371 379 POLY-SER.
 FT ZN_FING 538 560 C2HC-TYPE.
 FT DOMAIN 788 801 POLY-GLU.
 FT DOMAIN 989 995 POLY-GLU.
 FT DOMAIN 1019 1026 POLY-ARG.
 FT DOMAIN 1069 1078 POLY-GLU.
 FT DOMAIN 1147 1150 POLY-LYS.
 FT DOMAIN 1221 1242 GLU-RICH.
 FT DOMAIN 1267 1302 GLU-RICH.
 FT DOMAIN 1411 1414 POLY-GLU.
 FT DOMAIN 1593 1597 POLY-SER.
 FT DOMAIN 1643 1704 GLN/PRO-RICH.
 FT DOMAIN 1897 1977 MET-RICH.
 FT SITE 1546 1547 BREAKPOINT FOR TRANSLOCATION TO FORM
 FT MOZ-CBP.
 SQ SEQUENCE 2004 AA; 225054 MW; 9FFBAC3792854BA CRC64;

Query Match 4.4%; Score 175; DB 1; Length 2004;
 Best Local Similarity 22.7%; Pred. No. 0.26; Indels 88; Gaps 13;
 Matches 73; Conservative 48; Mismatches 113;

Qy 122 BEAPFTEPEQVPEAEFQNIIEAKEQIQLLHEMVHAEHVEGDLQQEDGPTGPQQ 181
 Db 1205 QESSETVEPKEDMF-----LPEERKEE-EEAQAEAEAEAGEEDAAASEVPAASPAD 1256
 Qy 182 EDDEFLMATVDVDFEILEPVSHEETHSVHVEETVSQDCNQCMEMMSQENPDSDSEP 241
 Db 1257 SSN-----SPETETKEPEVEEERKPRVSEEQRSQSEERQSELESPPEEEDAAAE 1308
 Qy 242 VVEDERLHDTDD-----VTQVVEEQAVYPLENE-GI----- 274
 Db 1309 AQND-----HADDEDDGHLESTKKKELEEQPTREDVKEEPGVQESFLDANMKREKIKD 1365
 Qy 275 -EITEVTAPPEDNPVEDSQVIVEVSIFPVVEEQEVPPTNRKTTDDPEQAKVKKKPKL 333
 Db 1366 KEETELDS-EEEQSHDTSVVVSEQMA-----GSEDDHEEDSHTKEELIEL 1409
 Qy 334 LNKEDTKIKELDAEAKLRKKGKIEEAVNAFVELVRKYPQSPRAYKQACEDDLAEKRR 393
 Db 1410 KEE-EEIHPHSLDL-----ETVQAVQSLTQEESEHEGAY--QDCBETLA----- 1451
 Qy 394 SNEVLRGAIETYQVASLPDVP 415
 Db 1452 -----ACQLQSYTQADEDP 1466

RESULT 12
 YNJI YEAST
 ID YNJI YEAST STANDARD; PRT; 1240 AA.
 AC P53935;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 141.5 kDa protein in YPT53-RHO2 intergenic region.
 GN YNL091W OR N2231.
 OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RA MEDLINE=96367601; PubMed=8771715;
 RX Garcia-Cantalejo J.M., Boskovic J., Jimenez A.;
 RT "Sequence analysis of a 14.2 kb fragment of Saccharomyces cerevisiae
 chromosome XIV that includes the ypt53, trnAla and gsr m2 genes and
 four new open reading frames.";
 RT Yeast 12:599-608(1996).
 RL Yeast 12:599-608(1996).
 CC -1- SIMILARITY: TO S.POMBE SPAC29E6.10C.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; X85811; CAAS9826.1; -;
 DR EMBL; Z71367; CAAS9567.1; -;
 DR PIR; S52734; S52734.
 DR Germline; 143097; -;
 DR SGD; S0005035; YNL091W.
 DR GO; GO:0009651; P:salinity response; IMP.
 KW Hypothetical protein.
 FT DOMAIN 756 761 POLY-GLU.
 SQ SEQUENCE 1240 AA; 141513 MW; 3FE9D265822D5778 CRC64;

Query Match 4.3%; Score 172.5; DB 1; Length 1240;
 Best Local Similarity 23.9%; Pred. No. 0.19;
 Matches 103; Conservative 66; Mismatches 175; Indels 87; Gaps 21;

Qy 144 EDAKEAQIQLLHEMVHAEHVE-GEGLQEDGPTGPQEDDFLMATVDVDFETLEPE 202
 Db 460 EDEEDYDD-----YSEVAEDSEEVSEYEGIEAVEKPEHDE-----KNGIRETLHLS 508
 Qy 203 VSH-----FTEHSYHVEETVSQDCNQCMEMMSQENP-----DSSEP 241
 Db 509 YDHDHKKQNHPPHHYHSTSTHSD-ELSEEYISDIELPHDPKHFRHDDDLGDDEDP 567
 Qy 242 VVEDERLHDTDDVTVQ--VVEQAVVEPLENEGEITE-----VTAPPEDNPVEDSQV- 293
 Db 568 EEDENEGDDEED-TYDSGLDTRLEEGKLIQIATKLLQSPRIMASYHEKQADNRK 626
 Qy 294 IVEVSIFPVVEEQEVPPTNRKTTDDPEQAKVKKKPKLKNPKDK-----TIKREL 345
 Db 627 LLOE-----LEEKKKRKEEKKKKRKEKKRKLQQLAKEEKKRKEEKKRKEEKKEL 681
 Qy 346 DAAEKLRK--RGKIEAVNAFVELVRKYPQSPRAYKQACEDDLAEKRR--SNEVLRG 400
 Db 682 EEREMRREARQKRVEEA-----KKKDEERKRLLEEQRREEMOEKQKKEELKKR 734
 Qy 401 AIETYQVASLPDVPADLLKLSLRKRRDRQQFLGHMRGSLTLQLR--VOLFPNDTSLK- 457
 Db 735 REEEKKIREQKLEQEKLOKE-KEEERQRLAE---DALRQKLENEQTSANILSAKP 790
 Qy 458 -NDLGVYLLIGNDNNAKYVEEVLVSTPNDFPAKHVGFILKAQNKIAESIPYLKEGIE 516
 Db 791 FTENGVGPNVSSQSHPNMTNYQEDNSCSINDE-----ILKMNVAASKPVSPGTFN 842
 Qy 517 SGD---PGTDD 524
 Db 843 VHDLLPSTNN 853

RESULT 13
 YHUE YEAST
 ID YHUE YEAST STANDARD; PRT; 465 AA.
 AC P38845;

DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical 51.1 kDa protein in DCD1-WRPL6 intergenic region.
 GN YHR146W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX MEDLINE=94378003; PubMed=8091229;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 RA Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
 RA Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
 RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
 RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
 RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
 RA Vaudin M.;
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT VIII.";
 RL Science 265:2077-2082(1994).
 CC -!- SIMILARITY: TO YEAST YN1173C.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U10397; AAB68982.1; -.
 DR PIR; S46759; S46759.
 DR Germline; 139464; -.
 DR SGD; S0001189; CRFL.
 DR GO; GO:0005634; C:nucleus; IC.
 DR GO; GO:0003677; F:DNA binding; IDA.
 KW Hypothetical protein.
 SQ SEQUENCE 465 AA; 5115 MW; 30880758F37991C7 CRC64;

Query Match 4.3%; Score 172; DB 1; Length 465;
 Best Local Similarity 22.8%; Pred. No. 0.06;
 Matches 90; Conservative 52; Mismatches 132; Indels 124; Gaps 16;
 QY 76 DLVDYBEVLGKGIYDADGDGDFDQDAKVLG-----LLCKPPRSAGPPTSNNKKNKRNKRSK 141
 DB 92 DLVETQEVAGASRIPEAG-----LLCKPPRSAGPPTSNNKKNKRNKRSK 141
 QY 109 LKERSTSEPAVPEAEPTPE-----EQVPEAEAPONIEDAKESQIQS 153
 DB 142 LKKKSTNNKKNKSNESLDDNEEDGVTGTTTDTGTSREETPL-ABPTNVSKEAPGNFHI 200
 QY 154 LLHEMVHAEHVEGEDIQQEDGPTGEP-----QQEDDFLMATDVRDFTLEPEVSHE 206
 DB 201 LPID-----QADTQSGIIGGGPVLVNPNGEIKETETRDVDAR-----E 243
 QY 207 ETEHSYHVEETVSQCNQDMNEMSEQEN--PDSSEPVVEDRLHDDTDVTVYVEEQ 264
 DB 244 LNERLNKEEVPPEVAGPIVSESVTEKSPALPOADDPVETKEVAHVQELTQV---EA 300
 QY 265 VYPELENEGIEITVAPPEDNPVDSQVIVEVS-IPFVEEQOEVPPETNKRKTDDEQ 323
 DB 301 V-TPLINE-----PPFLPTPEAQISIPSSKVEPVEGSLQ----- 334
 QY 324 AKVKKKPKLLNKFDKTIKAEKLRKRGKIEEAVNAFVKELVRKYPOSP-----RAR 378
 DB 335 -----SKLVEKRETS-EGVLGSKKVENKAKKDEEFTLDPIVNAKPLPLTDSQTAE 386
 QY 379 YGKAQCEDDLAEKRSRNEVLRGAIETVQEVASLPDVA 416
 DB 387 GRKSPAVSEKKEKKQEE--KGSKEVKRSKTSKEKKS 422

RESULT 14
 YCF2_OENV1
 ID YCF2_OENV1 STANDARD; PRT; 630 AA.
 AC P31569;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DE Hypothetical protein ycf2 (ORF 2280) (Fragment).
 GN YCF2.
 OS Oenothera villaricae.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC Myrtales; Onagraceae; Oenothera.
 OX NCBI_TaxID=3941;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93169690; PubMed=8435856;
 RA Nimyck R., Shoendorf T., Hachtel W.;
 RT "In-frame length mutations associated with short tandem repeats are
 RT located in unassigned open reading frames of Oenothera chloroplast
 RT DNA.";
 RL Curr. Genet. 23:265-270(1993).
 CC -!- SIMILARITY: Belongs to the ycf2 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X64615; CAA45896.1; -.
 DR PIR; S29796; S29796.
 KW Chloroplast; Hypothetical protein.
 FT NON TER 1
 SQ SEQUENCE 630 AA; 72781 MW; 6ABFFF7DC75B0BAA CRC64;
 Query Match 4.3%; Score 172; DB 1; Length 630;
 Best Local Similarity 20.9%; Pred. No. 0.087;
 Matches 107; Conservative 82; Mismatches 168; Indels 154; Gaps 22;
 QY 52 TSFFTFW-----MVALLGVTMSVAVVWFDLV-----DYEEVLKGLIYDADGDGD 97
 DB 100 SSIYKWFELGSMKKLTLLYLLTCSAGSIAQDLSPGPDEQNLTISYGLVENDSDLV 159
 QY 98 FDVDDAKVLLGLKERSTSEPAVPEAEPTPEEQVPEAEAPONIEDE---AKEIQSL 154
 DB 160 HGLSD--IVHGLELEGALVSGSPTEEEVEGTEEEVEGTEEEVEGTEEEVEGTE 217
 QY 155 LHEMVHA--EHVEG-----EDLQQEDGPTGEPQEDDFLMATDVRDFTLEPEV--SH 205
 DB 218 EEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTE 276
 QY 206 ETEHSYHVEETVSQCNQDMNEMSEQENPDSSEPVW---EDERLHDDTDVTVYVEEQ 256
 DB 277 EEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTE 336
 QY 257 -----YQVYE-----EQAVTEP--LENEGIRIT----- 277
 DB 337 PRNPLDIQRLIYHQKYESELEEDDDDDDDVPAQKMLEDFSELVWSPRIWHPWDFLLD 396
 QY 278 -EVTAPPENPVEDSQV---IVEEVSIPFVEE-----QQEVPPEPNRK 316
 DB 397 CRAEIPAEIPEEDELPEDALETEVAVWGVVEEAGEADEEDVLEAQOEDLLEEDSE 456
 QY 317 TDDPEQAKVKKKPKK---LLNKFDKTIKAEKLRKRGKIEEAVNAFVKELVRKYPOSP 351
 DB 457 LDBEDELDEEEEPKEEDELHHEEEEEEDEEDELQENDSEBFFRVKPIIPRHWTF 516

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 10:59:38 ; Search time 52 Seconds
(without alignments)
4599.285 Million cell updates/sec

Title: US-09-903-216-2

Perfect score: 4022

Sequence: 1 MAQRNNAKSSGSSSGSS.....IDVNHPELTPOQRSLPAI 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	4022	100.0	758	4	Q9Y4J0	Q9Y4J0 homo sapien
2	3209.5	79.8	739	11	Q9EPA6	Q9EPA6 mus musculus
3	3208.5	79.8	741	11	Q8BSY0	Q8BSY0 mus musculus
4	3175.5	79.0	725	11	Q8BQK0	Q8BQK0 mus musculus
5	2950	73.3	658	11	Q8CBM2	Q8CBM2 mus musculus
6	2919.5	72.6	689	11	Q9EQ66	Q9EQ66 mus musculus
7	1636	40.7	313	4	Q9H2C4	Q9H2C4 homo sapien
8	1457.5	36.2	299	4	Q9H291	Q9H291 homo sapien
9	1376.5	34.2	270	4	Q8TB28	Q8TB28 homo sapien
10	1066.5	26.5	785	5	Q9GQ82	Q9GQ82 drosophila
11	998.5	24.8	308	11	Q9EQ65	Q9EQ65 mus musculus
12	972.5	24.2	292	11	Q9IWG6	Q9IWG6 mus musculus
13	890	22.1	270	11	Q920F9	Q920F9 mus musculus
14	798	19.8	872	5	Q93178	Q93178 caenorhabdi
15	779.5	19.4	259	11	Q920F8	Q920F8 mus musculus
16	708.5	17.6	258	11	Q9EQ67	Q9EQ67 mus musculus

17	691.5	17.2	215	11	Q920F7	Q920F7 mus musculus
18	550.5	13.7	133	11	Q9EQ63	Q9EQ63 mus musculus
19	533.5	13.3	147	11	Q9EQ62	Q9EQ62 mus musculus
20	516	12.8	245	11	Q8CH79	Q8CH79 mus musculus
21	435.5	10.8	210	4	Q9NR11	Q9NR11 homo sapien
22	420.5	10.5	210	6	Q28264	Q28264 canis famil
23	418	10.4	225	4	Q9NR10	Q9NR10 homo sapien
24	407.5	10.1	207	11	Q9EQ64	Q9EQ64 mus musculus
25	407.5	10.1	207	11	Q9CR06	Q9CR06 mus musculus
26	407.5	10.1	212	11	Q9D7J8	Q9D7J8 mus musculus
27	390.5	9.7	304	2	Q53792	Q53792 streptomyce
28	276	6.9	343	11	Q80VF9	Q80VF9 mus musculus
29	275	6.8	250	16	Q93H17	Q93H17 streptomyce
30	274.5	6.8	343	4	Q9UH39	Q9UH39 homo sapien
31	263.5	6.6	186	4	Q9NSN3	Q9NSN3 homo sapien
32	257	6.4	176	6	Q9N1E7	Q9N1E7 oryctolagus
33	251	6.2	176	11	Q9CUZ2	Q9CUZ2 mus musculus
34	250	6.2	49	11	Q9EQ69	Q9EQ69 mus musculus
35	217.5	5.4	230	4	Q8N4H3	Q8N4H3 homo sapien
36	217.5	5.4	250	4	Q8N316	Q8N316 homo sapien
37	217.5	5.4	280	4	Q8IW63	Q8IW63 homo sapien
38	217.5	5.4	347	4	Q96H00	Q96H00 homo sapien
39	216.5	5.4	1110	13	Q91255	Q91255 petromyzon
40	207.5	5.2	1236	5	Q9GTX2	Q9GTX2 plasmodium
41	206	5.1	17352	5	Q95YM2	Q95YM2 procamburus
42	204.5	5.1	3111	5	Q9VH10	Q9VH10 drosophila
43	204	5.1	976	12	Q9DUN0	Q9DUN0 kaposi's sa
44	203	5.0	1233	5	Q8IJ56	Q8IJ56 plasmodium
45	199	4.9	1271	5	Q25860	Q25860 plasmodium

ALIGNMENTS

RESULT 1

Q9Y4J0 PRELIMINARY; PRT; 758 AA.

AC Q9Y4J0; DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Aspartyl(asparaginyl)beta-hydroxylase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96420598; PubMed=8823296;
RA Lavalisier L.; Jia S.; Nishiyama M., de la Monte S., Stern A.M.,
RA Wands J.R., Friedman P.A.;
RT "Overexpression of human aspartyl(asparaginyl)beta-hydroxylase in
RT hepatocellular carcinoma and cholangiocarcinoma.";
RL J. Clin. Invest. 98:1313-1323 (1996).
DR EMBL; S83325; AAB50779.1; -
DR InterPro; IPR007943; Asp-B-Hydro N.
DR InterPro; IPR007803; Asp Arg Hydros.
DR InterPro; IPR008940; Prenyl_trans.
DR InterPro; IPR001440; TPR.
DR Pfam; PF05279; Asp-B-Hydro N; 1.
DR Pfam; PF05118; Asp Arg_Hydro; 1.
SQ SEQUENCE 758 AA; 85890 MW; 4AF6F0AB4500AF0C CRC64;

Query Match 100.0%; Score 4022; DB 4; Length 758;
Best Local Similarity 100.0%; Pred. No. 9.1e-232;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAQRNNAKSSGSSSGSGSTSAGSSSPGARRTKGGHKNKRGKGLSGTSPFTFMV	60
Db	1	MAQRNNAKSSGSSSGSGSTSAGSSSPGARRTKGGHKNKRGKGLSGTSPFTFMV	60
QY	61	IALLGWTSVAVWEDLYEVLKLIYDADGGDDVDKVLGLKESTSEPAVP	120

Db 61 IALLGVWTSVAVVWFDLVYEEVGLKGIYDADGDGDFDVKAVLLGLKERSTSEPAVP 120
QY 121 PEAEAPTEPEEQVPVPAEAEQNIQSLHEMVAHVEGDIQOEDGPTGEPQ 180
Db 121 PEAEAPTEPEEQVPVPAEAEQNIQSLHEMVAHVEGDIQOEDGPTGEPQ 180
QY 181 QEDDEFMLATDVDDRFETLPEVSHETSHYVEETVSCQNDMEEMSEQENPDSSSE 240
Db 181 QEDDEFMLATDVDDRFETLPEVSHETSHYVEETVSCQNDMEEMSEQENPDSSSE 240
QY 241 PVVEDERLHDDTDVTVQVVEEQVAYEPELENEGIEITEVTAPPDNPVBSQVIVEVSI 300
Db 241 PVVEDERLHDDTDVTVQVVEEQVAYEPELENEGIEITEVTAPPDNPVBSQVIVEVSI 300
QY 301 FPVEEQVEVPEINRKTDDPEQAKVKKKPKLANKFKDKTIKABLDAAELKRGKLEEA 360
Db 301 FPVEEQVEVPEINRKTDDPEQAKVKKKPKLANKFKDKTIKABLDAAELKRGKLEEA 360
QY 361 VNAFKELVRYKYPQSPRARYKAQCEDDLAEKRRSNEVLRAIETVQVSLPDVADLLK 420
Db 361 VNAFKELVRYKYPQSPRARYKAQCEDDLAEKRRSNEVLRAIETVQVSLPDVADLLK 420
QY 421 LSLKRRSDRQOFLGHMGSLLTLQRLVQLPNDTSLKNDLGVGLLIGDNDNAKKVYEEV 480
Db 421 LSLKRRSDRQOFLGHMGSLLTLQRLVQLPNDTSLKNDLGVGLLIGDNDNAKKVYEEV 480
QY 481 LSVTPNDGFAKHYGFILKAQNKIAESIPLYKEGIESGDPGTDGGRFYFHLGDAMQVGN 540
Db 481 LSVTPNDGFAKHYGFILKAQNKIAESIPLYKEGIESGDPGTDGGRFYFHLGDAMQVGN 540
QY 541 KEAYKAYVELGHKRGHFAVWQSRSLYVNGLKAQPWTPKGTGYTELKYSERNWKLIRDE 600
Db 541 KEAYKAYVELGHKRGHFAVWQSRSLYVNGLKAQPWTPKGTGYTELKYSERNWKLIRDE 600
QY 601 GLAVMDKAKGLFLPEDENLREKGDWSOFTLWQOGRNENACKGAPKCTCTLLEKPEPTGC 660
Db 601 GLAVMDKAKGLFLPEDENLREKGDWSOFTLWQOGRNENACKGAPKCTCTLLEKPEPTGC 660
QY 661 RRQIKYISIMHPGTHVWPHGTPNCLRMHLGLVLPKEGCKIRCANETRTWEGKVLIFD 720
Db 661 RRQIKYISIMHPGTHVWPHGTPNCLRMHLGLVLPKEGCKIRCANETRTWEGKVLIFD 720
QY 721 DSPEHEWQDASSFRLLIFIVDVWHPHLPDPOORSLPAI 758
Db 721 DSPEHEWQDASSFRLLIFIVDVWHPHLPDPOORSLPAI 758

RESULT 2

Q9EPA6 ID Q9EPA6 PRELIMINARY; PRT; 739 AA.
AC Q9EPA6; PRELIMINARY; PRT; 739 AA.
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Aspartyl beta-hydroxylase 6.6 kb transcript (Aspartyl beta-hydroxylase
4.5 kb transcript).
GN 2310005F16RIK OR ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=20564328; PubMed=10956665;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved
RT Isoform of Asph Missing the Catalytic Domain Share Exons with
RT Junction".
RL J. Biol. Chem. 275:39543-39554 (2000).
DR EMBL; AF289487; AAC40809.1; -.

DR EMBL; AF289486; AAC40808.1; -.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-hydro N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF05279; Asp-B-Hydro N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
SQ SEQUENCE 739 AA; 82841 MW; 4DF9F642512CA4EB CRC64;

Query Match 79.8%; Score 3209.5; DB 11; Length 739;
Best Local Similarity 80.7%; Pred. No. 2.8e-183;
Matches 619; Conservative 41; Mismatches 70; Indels 37; Gaps 8;

QY 1 MAORKNAK-SSGNSSSSGSGSGS-----TSAGSSSPCARRETKHGKHKRGKGLSG 51
Db 1 MAPRKNAKGGGNSSSSGSGSGSPSTGSSSSSPGARREAKHGKHKRGKGGISG 60
QY 52 TSFTTFWMTALLGVWTSVAVVWFDLVYEEVGLKGIYDADGDGDFDVKAVLLGLKE 111
Db 61 GSFTTFWMTALLGVWTSVAVVWFDLVYEEVGLKGIYDADGDGDFDVKAVLLGLKE 120
QY 112 RSTSEPAVPEEAEPEHTEPEEQVPAEAEQNIQSLHEMVAHVEGDIQOEDGPTG 171
Db 121 RSPSERTFPP-EAEETHAELEEQAEQADIQNVDEVEKEIQSLIQESVHTDH----DL-E 174
QY 172 EDGPTGPEQOEDDEFMLATDVDDRFETLPEVSHETSHYVEETVSCQNDMEEMMS 231
Db 175 ADGLAGEPOEVEDFUTVDDSDRFDLPEFVHEEDTYHVEDTASQNHPNDEMETN 234
QY 232 EQENPDSSFPVDEDERLHDDTDVTVQVVEEQVAYEPELENEGIEITEVTAPPDNPVEDS 291
Db 235 EQENSDPSEAVTDAGVLLPHAEVVRHQDYDE-PVYEPSEHGVVIS-----DNTIDS 286
QY 292 QVTVEVSVTFPVEEQVEVPEINRKTDDPEQAKVKKKPKLANKFKDKTIKABLDAAELK 351
Db 287 SIIEEINVASVEEQDQTPP-----VKKKKPKLANKFKDKTIKABLDAAELK 332
QY 352 RKRGKLEEVAVNAPELVRYKYPQSPRARYKAQCEDDLAEKRRSNEVLRAIETVQV 411
Db 333 RKRGKLEEVAVNAPELVRYKYPQSPRARYKAQCEDDLAEKRRSNEVLRAIETVQV 392
QY 412 PDVPADLLKLSKRRSDRQOFLGHMGSLLTLQRLVQLPNDTSLKNDLGVGLLIGDND 471
Db 393 PDAPTLVKLSKRRSERQOFLGHMGSLLTLQRLVQLPNDTSLKNDLGVGLLIGDND 452
QY 472 NAKVYEEVSVTPNDGFAKHYGFILKAQNKIAESIPLYKEGIESGDPGTDGGRFYFHL 531
Db 453 SAKVYEEVSVTPNDGFAKHYGFILKAQNKIAESIPLYKEGIESGDPGTDGGRFYFHL 512
QY 532 GDAMQVGNKEAYKAYVELGHKRGHFAVWQSRSLYVNGLKAQPWTPKGTGYTELKYSLE 591
Db 513 GDAMQVGNKEAYKAYVELGHKRGHFAVWQSRSLYVNGLKAQPWTPKGTGYTELKYSLE 572
QY 592 RNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSOFTLWQOGRNENACKGAPKCTCTLL 651
Db 573 RNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSOFTLWQOGRNENACKGAPKCTCALL 632
QY 652 EKPEPTGCRGOIKYISIMHPGTHVWPHGTPNCLRMHLGLVLPKEGCKIRCANETRTW 711
Db 633 EKPEPTGCRGOIKYISIMHPGTHVWPHGTPNCLRMHLGLVLPKEGCKIRCANETRTW 692
QY 712 BEGKVLIFDPSPEHEWQDASSFRLLIFIVDVWHPHLPDPOORSLPAI 758
Db 693 BEGKVLIFDPSPEHEWQDASSFRLLIFIVDVWHPHLPDPOORSLPAI 739

RESULT 3

Q8BSY0 ID Q8BSY0 PRELIMINARY; PRT; 741 AA.
AC Q8BSY0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

```

DB 574 ERNWKLIIRDEGLAVMDKAKGLFLPEDENLRKGDWSQFTLWQQRKKNENACKGAPKTCAL 633
QY 651 LEKPEPTGCRGQIKYSIMHPTGTHVPHPTGTCNCRIMHGLIYIPKEGCKIRCANETRT 710
DB 634 LEKSETTGCRGQIKYSIMHPTGTHVPHPTGTCNCRIMHGLIYIPKEGCKIRCANETRT 693
QY 711 WEEGKVLIFDDSPFEHVEWQDASSFRLLFIVDVWHPHLPETPQORRSIPAI 758
DB 694 WEEGKVLIFDDSPFEHVEWQDASSFRLLFIVDVWHPHLPETPQORRSIPAI 741

RESULT 4
Q8BQKO PRELIMINARY; PRT; 725 AA.
AC Q8BQKO;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Aspartate-beta-hydroxylase.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.";
DR EMBL; AK049506; BAC33783.1; -.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-hydro N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF05279; Asp-B-Hydro N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
DR SMART; SM00028; TPR; 2.
SQ SEQUENCE 725 AA; 81492 MW; F3CE979F8FC9C3D5 CRC64;

Query Match 79.0%; Score 3175.5; DB 11; Length 725;
Best Local Similarity 79.8%; Pred. No. 2.9e-181;
Matches 613; Conservative 41; Mismatches 61; Indels 53; Gaps 9;

QY 1 MAQRKNAK-SGNSSSSGSGSGS-----TSAGSSSPGARRETKHGKHNKRGKGLSG 51
DB 1 MAPRKNKGGGNSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 60
QY 52 TSFTFWFWIALLGVTSAVWVFDLVYEEVGLKGIYDADGDGDFDVKVLLGLKE 111
DB 61 GSFTFWFWIALLGVTSAVWVFDLVYEEVGLKGIYDADGDGDFDVKVLLGLKE 120
QY 112 RSTSPVAPP-EEAPHTPEEQVPEAEAPONIEAEQIOSLHWMVHAHVGEGLQ 170
DB 121 RSPSERTFPPEEAETHAELEQAPEGADIQNVEDEVKEQIOSLQESVHTDH----DL- 175
QY 171 QEDGPTGPOQDDDFLMTDVRFTLEPEVSHHEEHSYHVEETVSQCNQDMEEKM 230
DB 176 EADGLAGEPQPEVEDFLVTDSDRFDLEPGTVHEEIEDTVHVEDTASQNHNDMEEMT 235
QY 231 SEQNPDSPEPVEDERLHDTDDVTYQVYEQAVYEPLENEGIEITVETAPPENPVED 290
DB 236 NEQEN-----SEVRHQDYDE-PYEPSEHGVAS-----DNTIDD 271
QY 291 SOVIVVEYSIFPVEEQVPEPETNKTDDPEQAKVKKKPKLKNKFKDKTIKAELDAAEK 350
DB 272 SSIISEENINVASVEQDTPP-----VKKKPKLKNKFKDKTIKAELDAAEK 317
QY 351 LKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRSNEVLRGAIETTYQEVAS 410

Query Match 79.8%; Score 3208.5; DB 11; Length 741;
Best Local Similarity 80.6%; Pred. No. 3.2e-183;
Matches 619; Conservative 41; Mismatches 71; Indels 37; Gaps 8;

QY 1 MAQRKNAK-SGNSSSSGSGSGS-----TSAGSSSPGARRETKHGKHNKRGKGLSG 51
DB 1 MAPRKNKGGGNSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 60
QY 52 TSFTFWFWIALLGVTSAVWVFDLVYEEVGLKGIYDADGDGDFDVKVLLGLKE 111
DB 61 GSFTFWFWIALLGVTSAVWVFDLVYEEVGLKGIYDADGDGDFDVKVLLGLKE 120
QY 112 RSTSPVAPP-EEAPHTPEEQVPEAEAPONIEAEQIOSLHWMVHAHVGEGLQ 170
DB 121 RSPSERTFPPEEAETHAELEQAPEGADIQNVEDEVKEQIOSLQESVHTDH----DL- 175
QY 171 QEDGPTGPOQDDDFLMTDVRFTLEPEVSHHEEHSYHVEETVSQCNQDMEEKM 230
DB 176 EADGLAGEPQPEVEDFLVTDSDRFDLEPGTVHEEIEDTVHVEDTASQNHNDMEEMT 235
QY 231 SEQNPDSPEPVEDERLHDTDDVTYQVYEQAVYEPLENEGIEITVETAPPENPVED 290
DB 236 NEQENSDPSEAVTDAGVLLPHAEVVRHQDYDE-PYEPSEHGVAS-----DNTIDD 287
QY 291 SOVIVVEYSIFPVEEQVPEPETNKTDDPEQAKVKKKPKLKNKFKDKTIKAELDAAEK 350
DB 288 SSIISEENINVASVEQDTPP-----VKKKPKLKNKFKDKTIKAELDAAEK 333
QY 351 LKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRSNEVLRGAIETTYQEVAS 410
DB 334 LKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRSNEVLRGAIETTYQEVAS 393
QY 411 LPDVPADLLKLSIKRRSDRQOFLGHWGSLITLQRLVQLFPNDTSLKNDLGVGYLLIGN 470
DB 394 LPDAPTDLKLKLSIKRRSERQQFLGHWGSLITLQRLVQLFPNDTSLKNDLGVGYLLIGN 453
QY 471 DNAKKVYEEVLSVTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGDDGGRFYFH 530
DB 454 DSAKKVYEEVLSVTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGDDGGRFYFH 513
QY 531 LGDAMQVGNKEAYKWYELGHRGHPASVWQSLYVNGLVKAPQWMTPTKETGYTELKSL 590
DB 514 LGDAMQVGNKEAYKWYELGHRGHPASVWQSLYVNGLVKAPQWMTPTKETGYTELKSL 573
QY 591 ERNWKLIIRDEGLAVMDKAKGLFLPEDENLRKGDWSQFTLWQQRKKNENACKGAPKTCAL 650
```

```

Db 318 LRKRGKIEEAVNAFEELVRKYPOSPRARYKCAQCEDDLAEKQSRNEVIRAEITYQEAAD 377
QY 411 LPDVPADLLKLSIKRRSDROQFLGHRGSLTTLQRLVQLFPNDTSLKNDLGVGYLLIGN 470
Db 378 LPDAPDLDVLSIKRRSERQOFLGHRGSLTTLQRLVQLFPNDTSLKNDLGVGYLLIGN 437
QY 471 DNAAKVVVEVLVTPNDGFAKVHYGFLKAKQNKIAESIPLYKEGIESGDPGTDGFRFYH 530
Db 438 DSAKKVVEVLVTPNDGFAKVHYGFLKAKQNKISESIPLYKEGIESGDPGTDGFRFYH 497
QY 531 LGDAMORVGNKEAYKMYELGHKGHGFASVWQSRSLYVNGLKAKQPMWTPRETGYTELKSL 590
Db 498 LGDAMORVGNKEAYKMYELGHKGHGFASVWQSRSLYVNGLKAKQPMWTPRETGYTELKSL 557
QY 591 ERNWKILRDEGLVMDKAKGLFLPEDENLREKGDWSOFTLWQGRNRNENACKGAPKTCIL 650
Db 558 ERNWKILRDEGLVMDKAKGLFLPEDENLREKGDWSOFTLWQGRNRNENACKGAPKTCAL 617
QY 651 LEKFETTCRRGQIKYISIMHPGTHVWPHGTPNCRLRMHLGLVLPKEGCKIRCANEETR 710
Db 618 LEKFETTCRRGQIKYISIMHPGTHVWPHGTPNCRLRMHLGLVLPKEGCKIRCANEETR 677
QY 711 WEEGKVLIFDDSEHEVWQDASSFRLLIFIVDVWHPELTPQORSLPAI 758
Db 678 WEEGKVLIFDDSEHEVWQDASSFRLLIFIVDVWHPELTPQORSLPAI 725

RESULT 5
Q8CEM2 PRELIMINARY; PRT; 658 AA.
AC Q8CEM2;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Aspartate-beta-hydroxylase.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Urinary bladder;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK035735; BAC29171.1; -.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-hydro N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR008940; Prenyl_trans.
DR InterPro; IPR001440; TPR.
DR Pfam; PF05279; Asp-B-Hydro N; 1.
DR Pfam; PF05118; Asp Arg Hydrox; 1.
DR SMART; SM00028; TPR; 2.
SQ SEQUENCE 658 AA; 75127 MW; 2BA0A5C8E06801C8 CRC64;

Query Match 73.3%; Score 2950; DB 11; Length 658;
Best local similarity 80.7%; Pred. No. 7.3e-168;
Matches 566; Conservative 38; Mismatches 53; Indels 44; Gaps 7;

QY 59 MVIALLGVTWSVAVWFDLVYDEVILGKLGITYADGDFDVEDDAKVLGLKERSISEPA 118
Db 1 MVIALLGVTWSVAVWFDLVYDEVILGKLGITYADGDFDVEDDAKVLGLKERSISEPT 60
QY 119 VPP-EEAEPTPEEQVFPVEAEQNIIDEAEKEIQISLLHVMVHAEHVEGEDLOEQDPTG 177
Db 61 FPPEEAEHAELEEQAPEGADIQNVEDEKIQISLLQESVHTDH---DL-EADGLAG 115

```

```

QY 178 EPQOEDDEFIMATDVEDTLEPEVSHHEETSHTVSDCNQDMSEVMSEQENPD 237
Db 116 EPQOVEDFELTVTDSDDREFDELPGVYHEIEDTYHVEDTASQNHNDMEEMNEQEN-- 173
QY 238 SSEPVVEDERLHHTDDVTYQVVEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEE 297
Db 174 -----SSEVRHQDYDE-PVYEPSEHEGVAIS-----DNTIDSSIISEE 211
QY 298 VSIFPVEEQOQVPPETNRKTDDEQAKYKVKKKPKLLNKFDTIKAELOAAEKLRKRGKI 357
Db 212 INVASVEEQOQTPP-----VKKKPKLLNKFDTIKAELOAAEKLRKRGKI 257
QY 358 EEAUNAFAKELVRKYPOSPRARYKCAQCEDDLAEKQSRNEVIRAEITYQEAADLPDPTD 417
Db 258 EEAUNAFAKELVRKYPOSPRARYKCAQCEDDLAEKQSRNEVIRAEITYQEAADLPDPTD 317
QY 418 LKLSIKRRSDROQFLGHRGSLTTLQRLVQLFPNDTSLKNDLGVGYLLIGNDNKVKY 477
Db 318 LVKLSIKRRSERQOFLGHRGSLTTLQRLVQLFPNDTSLKNDLGVGYLLIGNDNKVKY 377
QY 478 EVELVTPNDGFAKVHYGFLKAKQNKIAESIPLYKEGIESGDPGTDGFRFYHGLGDAMQR 537
Db 378 EVELVTPNDGFAKVHYGFLKAKQNKISESIPLYKEGIESGDPGTDGFRFYHGLGDAMQR 437
QY 538 VGNKEAYKMYELGHKGHGFASVWQSRSLYVNGLKAKQPMWTPRETGYTELKSLERNWKLI 597
Db 438 VGNKEAYKMYELGHKGHGFASVWQSRSLYVNGLKAKQPMWTPRETGYTELKSLERNWKLI 497
QY 598 RDEGLAVMDKAKGLFLPEDENLREKGDWSOFTLWQGRNRNENACKGAPKTCITLLEKFPET 657
Db 498 RDEGLAVMDKAKGLFLPEDENLREKGDWSOFTLWQGRNRNENACKGAPKTCITLLEKFPET 557
QY 658 TCRCRGQIKYISIMHPGTHVWPHGTPNCRLRMHLGLVLPKEGCKIRCANEETRWECKVL 717
Db 558 TCRCRGQIKYISIMHPGTHVWPHGTPNCRLRMHLGLVLPKEGCKIRCANEETRWECKVL 617
QY 718 IFDDSEHEVWQDASSFRLLIFIVDVWHPELTPQORSLPAI 758
Db 618 IFDDSEHEVWQDASSFRLLIFIVDVWHPELTPQORSLPAI 658

RESULT 6
Q9EQ66 PRELIMINARY; PRT; 689 AA.
AC Q9EQ66;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Aspartyl beta-hydroxylase.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=20564328; PubMed=10956665;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Pocht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RA "Aspartyl beta-hydroxylase (Asph) and an evolutionarily conserved
RT isoform of asph missing the catalytic domain share exons with
RT junction.";
RL J. Biol. Chem. 275:39543-39554(2000).
[2]
SEQUENCE FROM N.A.
RP Henderson N.L., Dinchuk J.E., Burn T.C., Hollis G.F., Friedman P.A.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF289215; AAG39913.1; -.
DR EMBL; AF289205; AAG39913.1; JOINED.
DR EMBL; AF289206; AAG39913.1; JOINED.
DR EMBL; AF289207; AAG39913.1; JOINED.
DR EMBL; AF289208; AAG39913.1; JOINED.

```

DR	ENBL; AF289209; AAG39913.1; JOINED.	
DR	ENBL; AF289210; AAG39913.1; JOINED.	
DR	ENBL; AF289211; AAG39913.1; JOINED.	
DR	ENBL; AF289212; AAG39913.1; JOINED.	
DR	ENBL; AF289213; AAG39913.1; JOINED.	
DR	ENBL; AF289214; AAG39913.1; JOINED.	
DR	MCD; MGI.1914186; Aaph.	
DR	InterPro; IPR007943; Asp-B-hydro N.	
DR	InterPro; IPR007803; Asp_Arg_HydroFox.	
DR	InterPro; IPR001440; TPR.	
DR	InterPro; IPR008941; TPR-like.	
DR	Pfam; PF05279; Asp-B-Hydro N; 1.	
DR	Pfam; PF05118; Asp_Arg_HydroFox; 1.	
DR	SEQUENCE 689 AA; 77319 MW; 9CB916DF109F432C CRC64;	

Query Match	72.6%;	Score 2919.5;	DB 11;	Length 689;
Best Local Similarity	74.7%;	Pred. No. 5.2e-166;		
Matches 573;	Conservative 39;	Mismatches 68;	Indels 87;	Gaps 9

QY	1	MAQRKNAK-SSGNSSSSGSGSGS-----TSAGSSSPGARRETKHGCHKNGRKGGLSG	51
DB	1	MAPRKNKAGGGNSSSSGSGSGSGSPSTGSSSSSGSGARR-----	43
QY	52	TSPTTFWFIALLGWTSVAVVWFVDLVYEBVLGKGIYDADGGDFDVKVLLGLKE	111
DB	44	-----GKLGVDADGGDFDVKVLLGLKE	70
QY	112	RSTSEPAVPPEAEPTHEPEQVPVEAEPQNIEDBAEQIOSLLHEMVAHHEVEGEDLQQ	171
DB	71	RSPERTFPP-EATHALEEQAEQADIQNVEDEVKEQIOSLQESVHTDH-----DI-E	124
QY	172	EDGPTGSPQSDDEFLMATVDVDDRPETLEPEVSHHEETSHYVEVTSQDNCQDMEEMWS	231
DB	125	ADGLIAGSPQPEVEDFLITVDSDDRFEDLEPGTVHBEIEDTVHVEDTASQNPNDMEEMTN	184
QY	232	EQENPDSEPEVVEDERLHHDITDVTYQVVEQAVVEPLENKGIEITETVATPEQNPVEDS	291
DB	185	EQENSDSEAVTDAGVLLPAAEEVRRHQDYDE-PVYEPSEHGEVLS-----DNTIDDS	236
QY	292	QVIVVEYSIFPVEQEQEVPPTNRKTDTPDQEKAKVKKKPKLLNKFDTIKAEIDAAEKL	351
DB	237	SIISEEINVASVEEQDTPP-----VKKKKPKLLNKFDTIKAEIDAAEKL	282
QY	352	RKRKGIIEAVNAFKELRYKYPQSPRARKCAQEDDLAEKRSNEVIRGAETTYQEVASL	411
DB	283	RKRKGIIEAVNAFRELVKYYPQSPRARKCAQEDDLAEKORSNEVIRRAETTYQEAADL	342
QY	412	PDVPADLLKLSKRSRDRQQFLGHMRGSLLLTLQRLVQLFPNDTSLKNDLGVYLLIGND	471
DB	343	PDAPTDLVKLSKRSRERQQFLGHMRGSLLLTLQRLVQLFPSPDTLLKNDLGVYLLIGND	402
QY	472	NAKKVVEVLSVTNDGPAKVHYGFILKAQNKIAESIPLYKEGIESGPGTDDGFRFYHLL	531
DB	403	SAKKVVEVLSVTNDGPAKVHYGFILKAQNKISIESIPLYKEGIESGPGTDDGFRFYHLL	462
QY	532	GDAMQRVGNKEAYKWYELGHRKGHFASVWQSRSLYNVNGLKAPWMTPKETGYTELVKSL	591
DB	463	GDAMQRVGNKEAYKWYELGHRKGHFASVWQSRSLYNVNGLKAPWMTPKETGYTELVKSL	522
QY	592	RNWKLIIRDEGLAVMDKAGLFLPDDENUREKGDWSQFTLWQGRNENACKGAPKTKTLL	651
DB	523	RNWKLIIRDEGLAVMDKAGLFLPDDENUREKGDWSQFTLWQGRNENACKGAPKTKTLL	582
QY	652	EKFPETTCRRGQIKYSIMHPGTHVPHPTNCLRMHLGLVIRPEKCKIRCANETRTW	711
DB	583	EKFPETTCRRGQIKYSIMHPGTHVPHPTNCLRMHLGLVIRPEKCKIRCANETRTW	642
QY	712	BEGKVLIPDDSFHEVHWODASSFRILFIVDVVHPELTPQQRSLPAI	758
DB	643	BEGKVLIPDDSFHEVHWODASSFRILFIVDVVHPELTPQQRSLPAI	689

RESULT 7

RT Chromosomal Localization, and Gene Structure of Junctate, a Novel
RT Integral Calcium Binding Protein of Sarco(endo)plasmic Reticulum
RL Membrane.";
RL J. Biol. Chem. 275:39555-39568 (2000).

DR EMBL; AF0306765; AAG42257.1; -

DR GO; GO:0005789; C:endooplasmic reticulum membrane; NAS.

DR GO; GO:0005509; F:calcium ion binding; NAS.

DR InterPro; IPR007943; Asp-B-hydro_N.

DR Pfam; PF05279; Asp-B-Hydro_N; 1.

SQ SEQUENCE 299 AA; 33815 MW; 658F88C34BC2CA37 CRC64;

Query Match

36.2%; Score 1457.5; DB 4; Length 299;

Best Local Similarity 94.2%; Pred. No. 3.1e-79;

Matches 278; Conservative 2; Mismatches 0; Indels 15; Gaps 1;

QY 34 RETKGGHNGKRGKGLSGTSFTFTWFMVIALLGWTSVAVVWFDLVDYERVL----- 84

DB 5 KETKGGHNGKRGKGLSGTSFTFTWFMVIALLGWTSVAVVWFDLVDYERVLAKAKDFRYN 64

QY 85 -----GKLGIVDADGDGDFDVKAKVLLGLKERSSTSEPAVPEEAPEHTEPEEQVPVEA 138

DB 65 LSEVLOGKLGIVDADGDGDFDVKAKVLLGLKERSSTSEPAVPEEAPEHTEPEEQVPVEA 124

QY 139 EPQNIIDEAKEQISLLHEMVHAEHVEGEDLQEDGPTGEPQEDDEFMLMATDVRPFT 198

DB 125 EPQNIIDEAKEQISLLHEMVHAEHVEGEDLQEDGPTGEPQEDDEFMLMATDVRPFT 184

QY 199 LEPEVSHETSHYHVEETVSDCNQDMSEMQENPDSPSEPVVDERLHHDHTDVTYQ 258

DB 185 LEPEVSHETSHYHVEETVSDCNQDMSEMQENPDSPSEPVVDERLHHDHTDVTYQ 244

QY 259 VYEEQAVYPLENEGIEITEVTAPPDNPVEDSQVIEEVSIPVBEQEQVPPET 313

DB 245 VYEEQAVYPLENEGIEITEVTAPPDNPVEDSQVIEEVSIPVBEQEQVPPET 299

RESULT 9

Q8TB28 PRELIMINARY; PRT; 270 AA.

ID Q8TB28

AC Q8TB28;

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Similar to aspartate beta-hydroxylase.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RN SEQUENCE FROM N.A.

RP TISSUE=Pancreas;

RA Strausberg R.;

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC025236; AH25236.1; -

DR InterPro; IPR007943; Asp-B-hydro_N.

DR Pfam; PF05279; Asp-B-Hydro_N; 1.

SQ SEQUENCE 270 AA; 29757 MW; 8551773C7272202A CRC64;

Query Match

34.28; Score 1376.5; DB 4; Length 270;

Best Local Similarity 85.94; Pred. No. 1.8e-74;

Matches 269; Conservative 1; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAQRKNAKSSGSSSGSGSGTSAGSSSPGARRETKHGHKNGKGLSGTSFTFMV 60

DB 1 MAQRKNAKSSGSSSGSGSGTSAGSSSPGARRETKHGHKNGKGLSGTSFTFMV 60

QY 61 IALLGWTSVAVVWFDLVDYEEVLGKLGIVDADGDGDFDVKAKVLLGLKERSSTSEPAVP 120

DB 61 IALLGWTSVAVVWFDLVDYEEVLGKLGIVDADGDGDFDVKAKVLLGLKERSSTSEPAVP 120

QY 121 PEEAPEHTEPEEQVPVEAEPQNIIDEAKEQISLLHEMVHAEHVEGEDLQEDGPTGEPQ 180

DB 121 PEEAPEHTEPEEQVPVEAEPQNIIDEAKEQISLLHEMVHAEHVEGEDLQEDGPTGEPQ 163

QY 181 QEDDEFMLMATDVRDFFETLEPVSHEETSHYHVEETVSDCNQDMSEMQENPDSPSE 240

DB 164 -----ETESHYHVEETVSDCNQDMSEMQENPDSPSE 197

QY 241 PVVEDERLHHDHTDVTYQVYEEQAVYPLENEGIEITEVTAPPDNPVEDSQVIEEVS 300

DB 198 PVVEDERLHHDHTDVTYQVYEEQAVYPLENEGIEITEVTAPPDNPVEDSQVIEEVS 257

QY 301 FPVEEQEQVPPET 313

DB 258 FPVEEQEQVPPET 270

RESULT 10

Q8QB28

ID Q8QB28

AC Q8QB28; PRELIMINARY; PRT; 785 AA.

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Aspartyl beta-hydroxylase variant 1 (CG8421-PA).

DE ASPH OR CG8421 OR CG18658.

GN Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RN SEQUENCE FROM N.A.

RX MEDLINE=20564328; PubMed=10956655;

RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,

RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,

RA Friedman P.A.;

RT "Aspartyl beta-hydroxylase (Asph) and an Evolutionarily Conserved

RT Isoform of Asph Missing the Catalytic Domain Share Exons with

RT Junctionin.";

RL J. Biol. Chem. 275:39543-39554 (2000).

RN [2]

RN SEQUENCE FROM N.A.

RC STRAIN=Berkeley;

RC MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fosler K.C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

QY	152	QSLHFWAHV	AHHVGGEDLQEDG-----PTGEQO-QEDDEFLMATVDVDR	199
DB	178	TA-----E	DAEEEDNDDEGTVATVATTAATTEATGETYEAEDEDEDAAD-DDA	231
QY	196	FETLEPEVS	-----HEHTEHSYHVETVSQDCHQDMEEMSEENPDSSEPVVVEDR	247
DB	232	VESTETAP	LUSKAEBOEDDEDEBEQIEESVBERGQYAAQAPAKNDDDD-----DDQ	288
QY	248	LHHD	DDVVYQVYBEQAVYPLENEGIEHTEVTAPPDNPVEDSQSVIVBEVSIFPVEEQ	307
DB	289	DKND	ADDDGDDDEPESLA--QQFEND-----PEDT-----VPAKAVESEEQD	328
QY	308	EVPP	ETNKTDDPE-----QKAKVKKPKLKNFKDKTIKAEKDAAEKLKRGKIEAVNAP	364
DB	329	QDADEE	PEKEGGSWLASQIKPEAAPAPAKED-PFEQELRKANEMIRENYAQAARSF	387
QY	365	KELVRK	PQSPRARYGAQCEDDLAEKRRSNEVLRGAIETYQVVASLDPVPADLLKLSLK	424
DB	388	NLT	TNFAHPSAHLGRALLELLAKKERSNQKLWEAIDAKRYLAFAGELVASNOEFQA	447
QY	425	RRS--DR	QFGLHMRGSLTLQRLVOLFPNDTSLKNDLGVGYLLIGDNDNAKKVVEEVL	482
DB	448	GESC	ENLRLFLGHRRQATTIHELLINLPEDPRLRNQLSITLYLWVNNLQQVEKVAETLK	507
QY	483	VT	PDNGPAKVHYGFIKA-QNKTAESIPYIKEGIESGDPDGDGRYFHLGDAMQVGNK	541
DB	508	LWPN	NAQLHYGLARQFHADYAKAPYLKYAVESGECTQEAFFYLSIGETMQRLSMK	567
QY	542	EAYK	WYELGHKRGHFASVWORSLYNVNGLKAOPWMTPKETGYTELKVSILERNKWLIR	600
DB	568	SEAL	EVGKGVAKGFFASLYQRSLYNPRUQAFWQPKETGYERQLEKLTINWRAIRO	627
QY	601	GLAV	MDKAKGLPELDENIREKGDWSQFTLWQQGRRENENACKGAPKTCITLLEKFPETTC	660
DB	628	GLALL	GRS-GFPEDEALLRDKGWQYQYELYAGRRVKONCRAPITCSLLEFPESAC	686
QY	661	RGQIK	YIMHPGTHVWPHGPTNCRULRMHLGLVLPK-EGCKIRCANETRTTWEEGKVLIF	719
DB	687	RRGQ	VKFSVMOAKTHVWPHGPTNCRULRAHLTLAAPEKASLURVAEQERTWREGELFIF	746
QY	720	DDSP	HEVWQDASSFRILFIVDVVHPBELTPQORRSLPAI	758
DB	747	DDSP	HEVWNGSOSRLVLIDMWHPOLSAQRRLSPI	785
RESULT	11			
Q9EQ65	ID	Q9EQ65	PRELIMINARY; PRT; 308 AA.	
AC	AC	Q9EQ65;		
DT	01-WAR-2001	(TREMBLrel. 16, Created)		
DT	01-WAR-2001	(TREMBLrel. 16, Last sequence update)		
DE	01-JUN-2003	(TREMBLrel. 24, Last annotation update)		
DE	Aspartyl beta-hydroxylase	2.8 kb transcript.		
GN	ASPH.			
OC	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
NCBI	TaxID=10090;			
RN	[1]_TaxID=10090;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BAUB/c; TISSUE=Liver;			
RC	MEDLINE=20564328; PubMed=10956665;			
RA	Dinchuk J.B., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,			
RA	O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,			
RA	Friedman P.A.;			
RT	"Aspartyl beta -hydroxylase (Asph) and an evolutionarily conserved			
RT	isoform of asph missing the catalytic domain share exons with			
RT	junctin.";			
RL	J. Biol. Chem. 275:39543-39554 (2000).			
DR	EMBL; AF289488; AAG40810.1; --.			
DR	MEDL; MGI:1914186; Asph.			
DR	InterPro; IPR007943; Asp-B-hydro N.			
DR	pfam; PF05279; Asp-B-Hydro N; 1.			

QY 1/2 EDGPTGEFQDEDEF LMAIDVDURFETLLSEPVSHSETEHSYHVEETVSQDCNQDMEEHMS 231

QC Eukaryota; Metazoa; Nematoda; Chromadorea; Knabalticida; Knabalticidae

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 11:01:09 ; Search time 23 Seconds
(without alignments)
1701.412 Million cell updates/sec

Title: US-09-903-216-2

Perfect score: 4022

Sequence: 1 MAQRKNAKSGNSGSSGSGS.....IVDVHPELTQRRSLPAI 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1334	33.2	255	3	US-09-040-485-2
2	1320	32.8	255	4	US-09-702-705-1806
3	1320	32.8	255	4	US-09-736-457-1806
4	1320	32.8	255	4	US-09-671-325-1806
5	1199	4.9	783	6	5231168-2
6	197.5	4.9	320	4	US-09-252-991A-29355
7	197.5	4.9	1018	1	US-08-072-610-2
8	197.5	4.9	1018	2	US-08-719-822B-2
9	197.5	4.9	1018	3	US-09-092-458-2
10	190.5	4.7	1162	4	US-08-728-323A-2
11	190.5	4.7	1162	4	US-09-298-568-2
12	189.5	4.7	1162	4	US-09-410-399-2
13	189.5	4.7	312	4	US-09-328-352-8015
14	177.5	4.4	565	3	US-08-961-083-218
15	177.5	4.4	565	3	US-09-536-784-218
16	177	4.4	327	4	US-09-489-039A-10158
17	172.5	4.3	700	4	US-09-107-532A-5094
18	163	4.1	411	2	US-08-741-134-6
19	163	4.1	2662	4	US-09-595-684B-31
20	162.5	4.0	310	4	US-09-252-991A-26511
21	162	4.0	1282	4	US-09-543-681A-5419
22	160	4.0	688	3	US-09-141-047-8
23	156.5	3.9	506	2	US-08-820-170A-19
24	156.5	3.9	506	3	US-09-055-699-19
25	156.5	3.9	506	3	US-09-273-565-19
26	156.5	3.9	506	4	US-09-565-538-19
27	156.5	3.9	506	4	US-09-661-468-19

28 156.5 3.9 506 4 US-09-976-165-19 Sequence 19, Appl
29 154.5 3.8 1196 4 US-09-107-532A-3944 Sequence 3944, Ap
30 153.5 3.8 1898 1 US-08-056-200-94 Sequence 94, Appl
31 153.5 3.8 1898 2 US-08-800-644-94 Sequence 94, Appl
32 152.5 3.8 984 1 US-08-242-932-2 Sequence 2, Appl
33 152.5 3.8 984 1 US-08-714-481-2 Sequence 2, Appl
34 152.5 3.8 984 5 PGT-US95-06111-2 Sequence 2, Appl
35 151.5 3.8 1972 4 US-08-875-435B-4 Sequence 4, Appl
36 151 3.8 404 4 US-09-554-080A-3 Sequence 3, Appl
37 151 3.8 436 4 US-09-554-080A-2 Sequence 3, Appl
38 150.5 3.7 3878 4 US-09-914-259-11 Sequence 11, Appl
39 150 3.7 571 4 US-09-216-393B-327 Sequence 327, App
40 149.5 3.7 1164 3 US-08-923-992A-10 Sequence 10, Appl
41 149 3.7 967 4 US-09-914-259-21 Sequence 21, Appl
42 148.5 3.7 816 2 US-08-533-306A-6 Sequence 6, Appl
43 148.5 3.7 816 2 US-08-742-923A-6 Sequence 6, Appl
44 148 3.7 258 3 US-08-961-083-90 Sequence 90, Appl
45 148 3.7 258 4 US-09-536-784-90 Sequence 90, Appl

ALIGNMENTS

RESULT 1
US-09-040-485-2
; Sequence 2, Application US/09040485
; Patent No. 6166176
; GENERAL INFORMATION:
; APPLICANT: Radosevich, James A.
; TITLE OF INVENTION: A GENE ENCODING A NOVEL MARKER FOR
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BRINKS, HOFER, GILSON & LIONE
; STREET: NBC tower - Suite 3600, 455 N. Cityfront
; STREET: Plaza Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60611-5599
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,485
; FILING DATE: 17-MAR-1998
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Martin, Alice O.
; REGISTRATION NUMBER: 35,601
; REFERENCE/DOCKET NUMBER: 8998/3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-321-4200
; TELEFAX: 312-321-4299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-040-485-2

Query Match 33.2%; Score 1334; DB 3; Length 255;
Best Local Similarity 99.6%; Pred. No. 1.6e-100;
Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 59 MYIALIGVWTSVAVVFDLVYEEVGLGIYDADGDGDFVDDAKVLLGLKERSTSEPA 118
Db 1 MYIALIGVWTSVAVVFDLVYEEVGLGIYDADGDGDFVDDAKVLLGLKERSTSEPA 60
QY 119 VPPEAEPTPEPEQVPVEAEFQNIQSLHEMVAEHVEGDLQOEDGPTGE 178

```
Db 61 VPPEEAPHTPEEQVPEAPQNIIEDEAKEQISLLHEMVAHVEGEDLQEDGPTGE 120
QY 179 PQOEDDEFLMATDVDDRFETLEPEVSHHEEHSYHVEETVSQDCNQDMEMMSQENPDS 238
Db 121 PQOEDDEFLMATDVDDRFETLEPEVSHHEEHSYHVEETVSQDCNQDMEMMSQENPDS 180
QY 239 SEPVEDERLHDDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPENPVEDSQVIVEEV 298
Db 181 SEPVEDERLHDDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPENPVEDSQVIVEEV 240
QY 299 SIFPVEEQEQVPPPT 313
Db 241 SIFPVEEQEQVPPPT 255

RESULT 2
US-09-702-705-1806
; Sequence 1806, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Lijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-702-705-1806

Query Match 32.8%; Score 1320; DB 4; Length 255;
Best Local Similarity 98.8%; Pred. No. 2.1e-99;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 59 MVTALLGWTSVAVWFDLVDEYVGLKGIYDADGDFDVEDDAKVLGLKERSTSEPA 118
Db 1 MVTALLGWTSVAVWFDLVDEYVGLKGIYDADGDFDVEDDAKVLGLKERSTSEPA 60
QY 119 VPPEEAPHTPEEQVPEAPQNIIEDEAKEQISLLHEMVAHVEGEDLQEDGPTGE 178
Db 61 VPPEEAPHTPEEQVPEAPQNIIEDEAKEQISLLHEMVAHVEGEDLQEDGPTGE 120
QY 179 PQOEDDEFLMATDVDDRFETLEPEVSHHEEHSYHVEETVSQDCNQDMEMMSQENPDS 238
Db 121 PQOEDDEFLMATDVDDRFETLEPEVSHHEEHSYHVEETVSQDCNQDMEMMSQENPDS 180
QY 239 SEPVEDERLHDDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPENPVEDSQVIVEEV 298
Db 181 SEPVEDERLHDDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPENPVEDSQVIVEEV 240
QY 299 SIFPVEEQEQVPPPT 313
Db 241 SIFPVEEQEQVPPPT 255

US-09-702-705-1806
```

```
RESULT 3
US-09-736-457-1806
; Sequence 1806, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Lijun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-736-457-1806

Query Match 32.8%; Score 1320; DB 4; Length 255;
Best Local Similarity 98.8%; Pred. No. 2.1e-99;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 59 MVTALLGWTSVAVWFDLVDEYVGLKGIYDADGDFDVEDDAKVLGLKERSTSEPA 118
Db 1 MVTALLGWTSVAVWFDLVDEYVGLKGIYDADGDFDVEDDAKVLGLKERSTSEPA 60
QY 119 VPPEEAPHTPEEQVPEAPQNIIEDEAKEQISLLHEMVAHVEGEDLQEDGPTGE 178
Db 61 VPPEEAPHTPEEQVPEAPQNIIEDEAKEQISLLHEMVAHVEGEDLQEDGPTGE 120
QY 179 PQOEDDEFLMATDVDDRFETLEPEVSHHEEHSYHVEETVSQDCNQDMEMMSQENPDS 238
Db 121 PQOEDDEFLMATDVDDRFETLEPEVSHHEEHSYHVEETVSQDCNQDMEMMSQENPDS 180
QY 239 SEPVEDERLHDDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPENPVEDSQVIVEEV 298
Db 181 SEPVEDERLHDDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPENPVEDSQVIVEEV 240
QY 299 SIFPVEEQEQVPPPT 313
Db 241 SIFPVEEQEQVPPPT 255

US-09-736-457-1806

RESULT 4
US-09-671-325-1806
; Sequence 1806, Application US/09671325
; Patent No. 6667154
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Lijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C12
; CURRENT APPLICATION NUMBER: US/09/671,325
; CURRENT FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 1825
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-09-671-325-1806

Query Match 32.8%; Score 1320; DB 4; Length 255;
Best Local Similarity 98.8%; Pred. No. 2.1e-99;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 59 MVIALLGWNTSVAVVDFLDVYEEVGLKGIYDADGDFDQDDAKVLLGLKERSTSEPA 118
Db 1 MVIALLGWNTSVAVVDFLDVYEEVGLKGIYDADGDFDQDDAKVLLGLKERSTSEPA 60

QY 119 VPPEAEHTEPEOVPVEAEAPONIEDEAKQIQLLHEMVAEHVGEDELOOEGPTGE 178
Db 61 VPPEAEHTEPEOVPVEAEAPONIEDEAKQIQLLHEMVAEHVGEDELOOEGPTGE 120

QY 179 PQOEDDEFLMATDVRDRETFLEPEVSHETESHVVEETVSDCQNDMEEMSEQENPDS 238
Db 121 PQOEDDEFLMATDVRDRETFLEPEVSHETESHVVEETVSDCQNDMEEMSEQENPDS 180

QY 239 SEPVEDERLHHDTDVTVQVYVEQAVPEPLNEGIEITEVTAPEDNPVEDSQVIVEEV 298
Db 181 SEPVEDERLHHDTDVTVQVYVEQAVPEPLNEGIEITEVTAPEDNPVEDSQVIVEEV 240

QY 299 SIFPVEEQOEVPPET 313
Db 241 SIFPVEEQOEVPPET 255

RESULT 5

Patent No. 5231168
; APPLICANT: DZIEGIEL, MORTEN, BORRE, MARTIN, JEPSEN, SOREN;
; YUUST, JENS, RIENECK, KLAUS, WIND, ANNETTE, JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO:2:
; LENGTH: 783
5231168-2

Query Match 4.9%; Score 199; DB 6; Length 783;
Best Local Similarity 20.6%; Pred. No. 1.6e-07;
Matches 143; Conservative 134; Mismatches 237; Indels 180; Gaps 36;

QY 81 BEVLGKLGIDYADGDFD-----DVDDAKVL---LGLKERSTSPAPVPE--BAE 125
Db 119 DEESGESGLVDNE-SGDPEEPNHEEFEPDQNDSELSENLVESEKSVSEPAHVIIVSEK 177

QY 126 PHTPEEQVPVEAEAPONIEDEAKQIQLLHEMVAEHVGEDELOOEGPTG----- 177
Db 178 SVSEPAEHVEIIVSEKST--SEPAEHVESV-----SEQSNNEPSEKKGFPVPSKPFEBIE 229

QY 178 ----EPQOEDDEFLMATDVRDRETFLEP-----EVSHEETESHVHEETVSDCQNCQ 224
Db 230 KVDVQPKIVDLQIIEPNFVDSQNPQEPVPEFVKIEKVPSEENKHA-SVDEPVEK--KE 286

QY 225 DMEEMMSEQENPDSE---PVVEERLHHDTDV--TYQVYEQAVVEPLENEGIEITEVT 280
Db 287 NVSEVVEEKQNSQESVEEIPVNEDE-----FEDVHTEQLDLHKTVDP---EIVEVEEIP 338

QY 281 APPENPNVEDSQVI--VEEVSIFP-----VEEQOEVPPETNRK 316
Db 339 SELHENEVAHPEIIVEEV--FPENQNNFQFIEDDDKSAHIQHEIIVEVEEILPEDDKN 396

QY 317 TDDPEQAKVKKKPKLLINKFKTKKAELDAAE-----KLKRGKIEEAVNAFKELVRK 370
Db 397 EKVEHEIIVEVEEILP-----DNKEGQHEIIVEVEEILPEDDKNEKVEHEIIVEVEEILPE 451

QY 371 YPQSPRARYGKACQEDDLAEKRRSNVLRGAIETTYQEVASLPDVPAD-----LLKLK 422
Db 452 -DKNEGQHEIIVEVEEILPEDK--NEKVEHEIIVEVEE-----LPEDKNEKQHEIIVEVE 503

QY 423 LKRRSDRQOFLGHMRGSLTLQLRVQLFPNDTSLKNDLG-----VGYLLIGDNDNAKV 476
Db 504 EILPEDKNEKQVH-----EIVEVEEILPED-----KNEKGQHEIIVEVEEILPEDKNEKQ 554

QY 477 YE--EVLVSVTNDGFAKHYGFI-----LKAQNKTAESIPYLKEGIESGDPGTD 523
Db 555 HEIIVEVEEILPEDKNEKQHEIIVEVEEILPEDKNEKQHEIIVEVEEILPEIVE----- 607

QY 524 DGRFYHGLDAMQVRGNKEAVKVELGKRGHFAFASVWORSLYN---VNLKAQPMWTPK- 579
Db 608 -----TEEVPSQNNNNENIETIKPEEKNEP-SVEKAIPQEPVPTLNENENWTPXP 659

QY 580 ---ETGYTELKVLERNWKLIRD-----EGLAVMDKAKGL--FLPEDENLREKGDWSQFT 629
Db 660 SEGESTKPDIVQI-----KIVQENKPKKETPVVDGPKHVEQNIQEDDNDDEDDDDIDFE 714

QY 630 LW-----QOGRNENACKGAPKTCITLLEKFFETT 658
Db 715 GLSRKXDEKSDSNKNKKKSSFTIYSTKKFKKVS 748

RESULT 6

US-09-252-991A-29355
; Sequence 29355, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29355
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29355

Query Match 4.9%; Score 197.5; DB 4; Length 320;
Best Local Similarity 25.9%; Pred. No. 5.7e-08;
Matches 58; Conservative 34; Mismatches 85; Indels 47; Gaps 9;

QY 551 HKRGHFAFASVWORSLYNVLNGLKAQPMWTPKTYGTYEL-----VKSL 590
Db 31 HFRGRARLPFLRQLVNHSA-----WFAP-----YNSLMYLFSSVPSPKPYLDRSRFFELDEL 81

QY 591 ERNWKILRDEGLAVMDKAKGLFLPEDENLRBK-----DWSQFTL-WQOGRRENACK 642
Db 82 KNNWQTIRREALNLFDG---YIRALANNNEAGFGSPFKGKWRFLYTWYDGL-PSAQ 137

QY 643 GAPTCTTLEKFPETTCGRRGQIKYSIMHPGTHVWHPGTPNCRIMHGLVIP-KEGCK 701
Db 138 LCPKTVELSVIRPNVKA-----MFTLLPGGSHLPHRDPFGGSLRYHLGLSTPNSDNCR 192

QY 702 IRCANETWTEGKVLIEDDSFEHEVWQDASSFRILFIVDVWHP 745
Db 193 IYVQGYAMRDGEDVMFDETFVHWKNETEQTRVILFCDIERP 236

RESULT 7

US-08-072-610-2
; Sequence 2, Application US/08072610
; Patent No. 5532133
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby and Darby
 ; STREET: 805 Third Ave.
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022-7513
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/072,610
 ; FILING DATE: 19930602
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gogoris, Adda
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 5986/07686
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212)527-7700
 ; TELEFAX: (212)753-6237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1018 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: YES
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: C-terminal
 ; ORIGINAL SOURCE:
 ; ORGANISM: Plasmodium vivax
 ; IMMEDIATE SOURCE:
 ; CLONE: pWMB3.3.1
 ; US-08-072-610-2

Query Match 4.9%; Score 197.5; DB 1; Length 1018;
 Best Local Similarity 23.2%; Pred. No. 3.1e-07;
 Matches 128; Conservative 86; Mismatches 205; Indels 133; Gaps 27;

QY 109 LKERSTSEPAVPEEAEPP-----HTEPEEQVPAEPQNIED-----EAKQIQSL 155
 DB 533 LKDPDAGEAVTPSKAPVQVPAVGAQVPTTELMQLQEDDFELEGTAAPESGELVL 592
 QY 156 HEMVHAHVHVEDLQOEDGP-TGEPOQED--DEFLMATDVRDFFETLEPEVSHETESHY 212
 DB 593 -----EGEGTEPEEPREGEPTGEVPEEELATPEDD-FELEEP--TGEVEETV 640
 QY 213 HVEETVSQDCNQDMEMMSQENPDSPVVEDERLHHDTHDDVTVQVYEEQAVVEPLENE 272
 DB 641 EGEETAEE--GEEVEEPAEVE-----EVEEVPAAVEEVEEVEEVEEVEEVP 681
 QY 273 GIEITEVAPPENPDVEDSIVIVVEYSIPFVEEQQVPPETNRKTDDEQKAKVKKKKPK 332
 DB 682 PAEVEEVEEVEE-----VEEVP-----EVEEVPAAVEEVEEVEEVEEVEEVP 728
 QY 333 LANKFDKTKIAELDAEAKLRKRGKIEEAVNAFKELVRKVPQSPRAYGKAQCDDDLAEKR 392
 DB 729 AVVEEVPAAVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 787
 QY 393 RS-----NEVLGATETVQEVASLPDVPADLLKLKLRSDRQQLGHRMGSLLTLQRL 446
 DB 788 LSIIDMHYQDVPKFEFMEEEETAVPLKPEDF-----AKEDSQSTEWLTFIQLEGDEWEL 843
 QY 447 VQLFPNDTSI-----KNDLGVGYLLIGNDNNAKKVYEVLSVTND--GFAKVHY 494
 DB 844 -----EVSUNKAREMFEQNKEMWAGWLRLIENKWE--YSQI-STKGDPAGLRKREW 894
 QY 495 G-----FTLKQNKIARESIPLYKEGIESGDGDTGDFYFLHGDAMQVRGNKEAYKY 547

DB 895 SDEKWKWFKFAEVKSQIDS---HLKKWMD-----THSNLFKILVKD-MSQFENKKTKEWL 946
 QY 548 -----ELGHKRGHF-----ASVWORSLYNVNGLKQAQ--PWWTPKRETY-- 583
 DB 947 MNHWKKNRGYSGSEFVMTTSLKLNVAKRSVWVRANPNINRRERELMKWFLKENEYLG 1006
 QY 584 --TELXSLERN 593
 DB 1007 QRMEKMDSLEKS 1018

RESULT 8
 US-08-719-822B-2
 ; Sequence 2, Application US/06719822B
 ; Patent No. 5874527
 ; GENERAL INFORMATION:
 ; APPLICANT: Barnwell, John
 ; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby and Darby
 ; STREET: 805 Third Ave.
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022-7513
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/719,822B
 ; FILING DATE: 09/30/96
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gogoris, Adda
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 5986/17686US2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212)527-7700
 ; TELEFAX: (212)753-6237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1018 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: YES
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: C-terminal
 ; ORIGINAL SOURCE:
 ; ORGANISM: Plasmodium vivax
 ; IMMEDIATE SOURCE:
 ; CLONE: pWMB3.3.1
 ; US-08-719-822B-2

Query Match 4.9%; Score 197.5; DB 2; Length 1018;
 Best Local Similarity 23.2%; Pred. No. 3.1e-07;
 Matches 128; Conservative 86; Mismatches 205; Indels 133; Gaps 27;

QY 109 LKERSTSEPAVPEEAEPP-----HTEPEEQVPAEPQNIED-----EAKQIQSL 155
 DB 533 LKDPDAGEAVTPSKAPVQVPAVGAQVPTTELMQLQEDDFELEGTAAPESGELVL 592
 QY 156 HEMVHAHVHVEDLQOEDGP-TGEPOQED--DEFLMATDVRDFFETLEPEVSHETESHY 212
 DB 593 -----EGEGTEPEEPREGEPTGEVPEEELATPEDD-FELEEP--TGEVEETV 640
 QY 213 HVEETVSQDCNQDMEMMSQENPDSPVVEDERLHHDTHDDVTVQVYEEQAVVEPLENE 272
 DB 641 EGEETAEE--GEEVEEPAEVE-----EVEEVPAAVEEVEEVEEVEEVEEVP 681

[illegible]


```
;
;
;   ZIP: 20850
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;   COMPUTER: HP Vectra 486/33
;   OPERATING SYSTEM: MSDOS version 6.2
;   SOFTWARE: ASCII Text
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/536,784
;   FILING DATE: 30-Oct-1997
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/961,083
;   FILING DATE: OCT-30-1997
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Michelle S. Marks
;   REGISTRATION NUMBER: 41,971
;   REFERENCE/DOCKET NUMBER: PB340P3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (301) 309-8504
;   TELEFAX: (301) 309-8512
;   INFORMATION FOR SEQ ID NO: 218:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 565 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLSCULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 218:
;
US-09-536-784-218
```

```
Query Match      4.4%; Score 177.5; DB 4; Length 565;
Best Local Similarity 22.5%; Pred. No. 5.5e-06;
Matches 114; Conservative 71; Mismatches 199; Indels 123; Gaps 22;

QY 111 ERSTSEPAVPPPE--APHTPEBEQVPEVPAEPQNIEDAEKEQIQSLLENVHAEHVHGED 168
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 169 LQEDGPTGPQOQDEFLMATVDORFETLEPEVSHSEETESHVVEYTSQDCNQ-DME 227
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 EKAPVPEKQPEAPEE--KAVETPKQESTFTKAEETVEP--KEETVNQSIQEPKVE 172
QY 228 EMNSEQNPDSPEPVVEDE----RLHDTDDVTVYVEQAVYEPLENEGIEITEVTAPP 283
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 173 TPAVEKQTEPTEPEKVEQAGEPVAPREDEQAPTAPVEPEKQPEVPEBEKAV---EETPKP 229
QY 284 ED-----NPVEDSQV---IVEVSIFFVE-----EQ 307
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 230 EDKIKGIGTKPEVDKSELNNQIDKASSVPTDYSTASYNALGPVLETAKGYVASEPVKQP 289
QY 308 EVPPETNR-----KTD-----DPEQAKAKKKPKLNLKPKDKTIKAELEDAEK 350
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 290 EVNSEINKLKTALDNLVNDKTELNNTIADAKTKVKEHYSDRSWQNLQTEVTKAEKVAANT 349
QY 351 LRKRGKIEEAVNAF-----KELVRKYPOSPPRARYGKAQ 383
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 384 CEDDLAEKRRSNEVLRGAIETYQVASLPDVPADLLKLKRRSDRQOFLGHMREGSLTL 443
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 407 IKSITAEKKGEEVINTVLTDDKVTI-ETISAAFKNLEYKYVTLSTTMIYDRNGEET 465
QY 444 QRL----VQLFPNDTSLKN----DLGVGYLLIGDNDNAKVVYEWLSVTPNDGPAKVHYG 495
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 466 ETLENQNIQLDLKKVELKNIKRTDL-INY-----ENGETNESLIITIPDD---KSNYY 515
QY 496 FILKAQNKIAB--SIPYLKEGIESGDP 520
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 516 LKITSNNQKTTLLAVKNIEETTNGTP 542
```

Search completed: May 5, 2004, 11:05:05
Job time : 25 secs


```

QY 181 QEDDEFLMATDVDDRETTLEPEVSHBETESHYVEETVSQDCNQDMEMMSBOENPDSSE 240
DB 181 QEDDEFLMATDVDDRETTLEPEVSHBETESHYVEETVSQDCNQDMEMMSBOENPDSSE 240
QY 241 PVVEDERLHDDTDDVTVQVVEQAVVEPLENEGIEITEVTAPEDNPVEDSQVIVEVSI 300
DB 241 PVVEDERLHDDTDDVTVQVVEQAVVEPLENEGIEITEVTAPEDNPVEDSQVIVEVSI 300
QY 301 FPVEEQEVPETNRKTDDPEQAKVKKPKLLNFKDKTKIAELDAAEKLKRGKIEEA 360
DB 301 FPVEEQEVPETNRKTDDPEQAKVKKPKLLNFKDKTKIAELDAAEKLKRGKIEEA 360
QY 361 VNAFKELVRKYPOSPRARYKQACEDDLAEKRSNEVLRGAIETYQEVASLPDVPADLLK 420
DB 361 VNAFKELVRKYPOSPRARYKQACEDDLAEKRSNEVLRGAIETYQEVASLPDVPADLLK 420
QY 421 LSLKRRSDRQOQFLGHRGSLTTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKYYEEV 480
DB 421 LSLKRRSDRQOQFLGHRGSLTTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKYYEEV 480
QY 481 LSVTPNDGFAKHYGFTLKQAKVKKPKLLNFKDKTKIAELDAAEKLKRGKIEEA 540
DB 481 LSVTPNDGFAKHYGFTLKQAKVKKPKLLNFKDKTKIAELDAAEKLKRGKIEEA 540
QY 541 KEAYKVELGHKRGHPASVWORSIYNVNGIKAPWMTPKETGYTELKSLERXWKLIRDE 600
DB 541 KEAYKVELGHKRGHPASVWORSIYNVNGIKAPWMTPKETGYTELKSLERXWKLIRDE 600
QY 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQOGRRENENACKGAPKCTTILEKFFETTC 660
DB 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQOGRRENENACKGAPKCTTILEKFFETTC 660
QY 661 RRGQIKYSIMHPGTHVWPHGTNCRLRMHLGLVIPKEGCKIRCANETRWEESKVLIFD 720
DB 661 RRGQIKYSIMHPGTHVWPHGTNCRLRMHLGLVIPKEGCKIRCANETRWEESKVLIFD 720
QY 721 DSFEHEVWQDASSFRLLIFIVDVVWHPELTPQORRSLPAI 758
DB 721 DSFEHEVWQDASSFRLLIFIVDVVWHPELTPQORRSLPAI 758

RESULT 2
US-09-859-604-2
; Sequence 2, Application US/09859604
; Patent No. US2002011059A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: Deutch, Alan H
; APPLICANT: Ghanbari, Hossein A
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 CIP
; CURRENT APPLICATION NUMBER: US/09/859,604
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-859-604-2

Query Match 100.0%; Score 4022; DB 9; Length 758;
Best Local Similarity 100.0%; Pred. No. 1.5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQRNAKSSGSSGSGSGSTAGSSPGARRETHKGHGKNGKGLSGTSFPTFMV 60
DB 1 MAQRNAKSSGSSGSGSGSGSTAGSSPGARRETHKGHGKNGKGLSGTSFPTFMV 60

```

```

QY 61 IALLGWTSVAVVWFEDLVYEEVLKGIYDADGDDGDFVDDAKVLLGLKERSTSPAYP 120
DB 61 IALLGWTSVAVVWFEDLVYEEVLKGIYDADGDDGDFVDDAKVLLGLKERSTSPAYP 120
QY 121 PEEAEPHTPEEQVPVEAEPQNIIEDEAKEQI QSLLEHMYHABHVEGEDIQQEDGPTGEQ 180
DB 121 PEEAEPHTPEEQVPVEAEPQNIIEDEAKEQI QSLLEHMYHABHVEGEDIQQEDGPTGEQ 180
QY 181 QEDDEFLMATDVDDRETTLEPEVSHBETESHYVEETVSQDCNQDMEMMSBOENPDSSE 240
DB 181 QEDDEFLMATDVDDRETTLEPEVSHBETESHYVEETVSQDCNQDMEMMSBOENPDSSE 240
QY 241 PVVEDERLHDDTDDVTVQVVEQAVVEPLENEGIEITEVTAPEDNPVEDSQVIVEVSI 300
DB 241 PVVEDERLHDDTDDVTVQVVEQAVVEPLENEGIEITEVTAPEDNPVEDSQVIVEVSI 300
QY 301 FPVEEQEVPETNRKTDDPEQAKVKKPKLLNFKDKTKIAELDAAEKLKRGKIEEA 360
DB 301 FPVEEQEVPETNRKTDDPEQAKVKKPKLLNFKDKTKIAELDAAEKLKRGKIEEA 360
QY 361 VNAFKELVRKYPOSPRARYKQACEDDLAEKRSNEVLRGAIETYQEVASLPDVPADLLK 420
DB 361 VNAFKELVRKYPOSPRARYKQACEDDLAEKRSNEVLRGAIETYQEVASLPDVPADLLK 420
QY 421 LSLKRRSDRQOQFLGHRGSLTTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKYYEEV 480
DB 421 LSLKRRSDRQOQFLGHRGSLTTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKYYEEV 480
QY 481 LSVTPNDGFAKHYGFTLKQAKVKKPKLLNFKDKTKIAELDAAEKLKRGKIEEA 540
DB 481 LSVTPNDGFAKHYGFTLKQAKVKKPKLLNFKDKTKIAELDAAEKLKRGKIEEA 540
QY 541 KEAYKVELGHKRGHPASVWORSIYNVNGIKAPWMTPKETGYTELKSLERXWKLIRDE 600
DB 541 KEAYKVELGHKRGHPASVWORSIYNVNGIKAPWMTPKETGYTELKSLERXWKLIRDE 600
QY 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQOGRRENENACKGAPKCTTILEKFFETTC 660
DB 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQOGRRENENACKGAPKCTTILEKFFETTC 660
QY 661 RRGQIKYSIMHPGTHVWPHGTNCRLRMHLGLVIPKEGCKIRCANETRWEESKVLIFD 720
DB 661 RRGQIKYSIMHPGTHVWPHGTNCRLRMHLGLVIPKEGCKIRCANETRWEESKVLIFD 720
QY 721 DSFEHEVWQDASSFRLLIFIVDVVWHPELTPQORRSLPAI 758
DB 721 DSFEHEVWQDASSFRLLIFIVDVVWHPELTPQORRSLPAI 758

RESULT 3
US-09-903-063-2
; Sequence 2, Application US/09903063
; Patent No. US20020114810A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV3
; CURRENT APPLICATION NUMBER: US/09/903,063
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-063-2

Query Match 100.0%; Score 4022; DB 9; Length 758;

```



```
; FILE REFERENCE: 21486-032 DIV4
; CURRENT APPLICATION NUMBER: US/09/903,199
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-199-2

Query Match      100.0%; Score 4022; DB 9; Length 758;
Best Local Similarity 100.0%; Pred. No. 1.5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAQRKNAKSSGSSSSGSGSTAGSSSPGARRETKHGKHGKRGKGLSGTSFPTWFMV 60
Db 1 MAQRKNAKSSGSSSSGSGSTAGSSSPGARRETKHGKHGKRGKGLSGTSFPTWFMV 60
Qy 61 IALLGVWTSVAVVWFDLVYEEVLGKGIYDADGDGDFVDDAKVLLGLKERSTSEPAVP 120
Db 61 IALLGVWTSVAVVWFDLVYEEVLGKGIYDADGDGDFVDDAKVLLGLKERSTSEPAVP 120
Qy 121 PEEAEPHTEPEEQVPEAEQNIIDEAKEQIQSLLEHMHVHAHVHVEGDLQEDGPTGEPQ 180
Db 121 PEEAEPHTEPEEQVPEAEQNIIDEAKEQIQSLLEHMHVHAHVHVEGDLQEDGPTGEPQ 180
Qy 181 QEDDEFMATDVDDRFETLEPEVSHETEHSHYHVEETVSQCNQDMSEMQENPNDSSE 240
Db 181 QEDDEFMATDVDDRFETLEPEVSHETEHSHYHVEETVSQCNQDMSEMQENPNDSSE 240
Qy 241 PVVEDERLHDDTDVTVQVVEEQAVYEPLENEGIEITETVAPPEDNPVEDSQVIVEVSI 300
Db 241 PVVEDERLHDDTDVTVQVVEEQAVYEPLENEGIEITETVAPPEDNPVEDSQVIVEVSI 300
Qy 301 FPVEEQEVPPETNRKTDDEPEQAKVKKKPKLANKFKDKTIKAELEDAEAKLRKKGKIEEA 360
Db 301 FPVEEQEVPPETNRKTDDEPEQAKVKKKPKLANKFKDKTIKAELEDAEAKLRKKGKIEEA 360
Qy 361 VNAFELVRKYPOSPRARYGKACEDDLAEKRSNEVLRGAEITYQVVASLPDVPADLLK 420
Db 361 VNAFELVRKYPOSPRARYGKACEDDLAEKRSNEVLRGAEITYQVVASLPDVPADLLK 420
Qy 421 LSLKRRSDRQOQFLGHMRGSLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKVVYEEV 480
Db 421 LSLKRRSDRQOQFLGHMRGSLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKVVYEEV 480
Qy 481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDGGRFYFHLGDAMQVRGN 540
Db 481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDGGRFYFHLGDAMQVRGN 540
Qy 541 KEAYKMYELGHRGHFASVWQSRSLYNNGLKAPQWMTPEKTYTELKSLERNKLLIRDE 600
Db 541 KEAYKMYELGHRGHFASVWQSRSLYNNGLKAPQWMTPEKTYTELKSLERNKLLIRDE 600
Qy 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRNENACKGAPKCTTLEKEPETTGC 660
Db 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRNENACKGAPKCTTLEKEPETTGC 660
Qy 661 RRGQIKYSIMHPGTHVWPHGTPTNCRMLRMLGLVLPKSGCKIRCANETRTWEEGKVLIFD 720
Db 661 RRGQIKYSIMHPGTHVWPHGTPTNCRMLRMLGLVLPKSGCKIRCANETRTWEEGKVLIFD 720
Qy 721 DSFEHEVWQDASSPRLIFIVDVVWHPBELTPQORRSLPAI 758
Db 721 DSFEHEVWQDASSPRLIFIVDVVWHPBELTPQORRSLPAI 758
```

RESULT 6

US-09-903-023-2

; Sequence 2, Application US/09903023

```
; Patent No. US20020146421A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV1
; CURRENT APPLICATION NUMBER: US/09/903,023
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-023-2

Query Match      100.0%; Score 4022; DB 9; Length 758;
Best Local Similarity 100.0%; Pred. No. 1.5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAQRKNAKSSGSSSSGSGSTAGSSSPGARRETKHGKHGKRGKGLSGTSFPTWFMV 60
Db 1 MAQRKNAKSSGSSSSGSGSTAGSSSPGARRETKHGKHGKRGKGLSGTSFPTWFMV 60
Qy 61 IALLGVWTSVAVVWFDLVYEEVLGKGIYDADGDGDFVDDAKVLLGLKERSTSEPAVP 120
Db 61 IALLGVWTSVAVVWFDLVYEEVLGKGIYDADGDGDFVDDAKVLLGLKERSTSEPAVP 120
Qy 121 PEEAEPHTEPEEQVPEAEQNIIDEAKEQIQSLLEHMHVHAHVHVEGDLQEDGPTGEPQ 180
Db 121 PEEAEPHTEPEEQVPEAEQNIIDEAKEQIQSLLEHMHVHAHVHVEGDLQEDGPTGEPQ 180
Qy 181 QEDDEFMATDVDDRFETLEPEVSHETEHSHYHVEETVSQCNQDMSEMQENPNDSSE 240
Db 181 QEDDEFMATDVDDRFETLEPEVSHETEHSHYHVEETVSQCNQDMSEMQENPNDSSE 240
Qy 241 PVVEDERLHDDTDVTVQVVEEQAVYEPLENEGIEITETVAPPEDNPVEDSQVIVEVSI 300
Db 241 PVVEDERLHDDTDVTVQVVEEQAVYEPLENEGIEITETVAPPEDNPVEDSQVIVEVSI 300
Qy 301 FPVEEQEVPPETNRKTDDEPEQAKVKKKPKLANKFKDKTIKAELEDAEAKLRKKGKIEEA 360
Db 301 FPVEEQEVPPETNRKTDDEPEQAKVKKKPKLANKFKDKTIKAELEDAEAKLRKKGKIEEA 360
Qy 361 VNAFELVRKYPOSPRARYGKACEDDLAEKRSNEVLRGAEITYQVVASLPDVPADLLK 420
Db 361 VNAFELVRKYPOSPRARYGKACEDDLAEKRSNEVLRGAEITYQVVASLPDVPADLLK 420
Qy 421 LSLKRRSDRQOQFLGHMRGSLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKVVYEEV 480
Db 421 LSLKRRSDRQOQFLGHMRGSLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKVVYEEV 480
Qy 481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDGGRFYFHLGDAMQVRGN 540
Db 481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDGGRFYFHLGDAMQVRGN 540
Qy 541 KEAYKMYELGHRGHFASVWQSRSLYNNGLKAPQWMTPEKTYTELKSLERNKLLIRDE 600
Db 541 KEAYKMYELGHRGHFASVWQSRSLYNNGLKAPQWMTPEKTYTELKSLERNKLLIRDE 600
Qy 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRNENACKGAPKCTTLEKEPETTGC 660
Db 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRNENACKGAPKCTTLEKEPETTGC 660
Qy 661 RRGQIKYSIMHPGTHVWPHGTPTNCRMLRMLGLVLPKSGCKIRCANETRTWEEGKVLIFD 720
Db 661 RRGQIKYSIMHPGTHVWPHGTPTNCRMLRMLGLVLPKSGCKIRCANETRTWEEGKVLIFD 720
Qy 721 DSFEHEVWQDASSPRLIFIVDVVWHPBELTPQORRSLPAI 758
```



```

; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-902-941-1806

Query Match 32.8%; Score 1320; DB 9; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 59 MVIALLGWTSVAVVWFDLVYEEVLGKGIYDADGDFDVEDDAAKVLGLKERSTSEPA 118
Db 1 MVIALLGWTSVAVVWFDLVYEEVLGKGIYDADGDFDVEDDAAKVLGLKERSTSEPA 60

QY 119 VPPEEAPHTPEEQVPVEAEAPONIEDEAKEQIQLLHMHVHAHVHVEGEDLQOEDGPTGE 178
Db 61 VPPEEAPHTPEEQVPVEAEAPONIEDEAKEQIQLLHMHVHAHVHVEGEDLQOEDGPTGE 120

QY 179 PQOEDDEFLMATDVRDFFETLEPEVSHEETSHTSHVVEETVSQDCNQDMMEEMSEQENPDS 238
Db 121 PQOEDDEFLMATDVRDFFETLEPEVSHEETSHTSHVVEETVSQDCNQDMMEEMSEQENPDS 180

QY 239 SEPVEDERLHHTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
Db 181 SEPVEDERLHHTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 240

QY 299 SIFPVEEQEVPDET 313
Db 241 SIFPVEEQEVPDET 255

RESULT 11
US-10-283-017-1806
; Sequence 1806, Application US/10283017
; Publication No. US200302111510A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Kaleo, Michael D.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Durham, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C20
; CURRENT APPLICATION NUMBER: US/10/283,017
; CURRENT FILING DATE: 2002-10-28
; NUMBER OF SEQ ID NOS: 2157
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-283-017-1806

Query Match 32.8%; Score 1320; DB 12; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 59 MVIALLGWTSVAVVWFDLVYEEVLGKGIYDADGDFDVEDDAAKVLGLKERSTSEPA 118
Db 1 MVIALLGWTSVAVVWFDLVYEEVLGKGIYDADGDFDVEDDAAKVLGLKERSTSEPA 60

QY 119 VPPEEAPHTPEEQVPVEAEAPONIEDEAKEQIQLLHMHVHAHVHVEGEDLQOEDGPTGE 178
Db 61 VPPEEAPHTPEEQVPVEAEAPONIEDEAKEQIQLLHMHVHAHVHVEGEDLQOEDGPTGE 120

QY 179 PQOEDDEFLMATDVRDFFETLEPEVSHEETSHTSHVVEETVSQDCNQDMMEEMSEQENPDS 238
Db 121 PQOEDDEFLMATDVRDFFETLEPEVSHEETSHTSHVVEETVSQDCNQDMMEEMSEQENPDS 180

QY 239 SEPVEDERLHHTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
Db 181 SEPVEDERLHHTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 240

QY 299 SIFPVEEQEVPDET 313
Db 241 SIFPVEEQEVPDET 255

```

```

; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-902-941-1806

Query Match 32.8%; Score 1320; DB 9; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 59 MVIALLGWTSVAVVWFDLVYEEVLGKGIYDADGDFDVEDDAAKVLGLKERSTSEPA 118
Db 1 MVIALLGWTSVAVVWFDLVYEEVLGKGIYDADGDFDVEDDAAKVLGLKERSTSEPA 60

QY 119 VPPEEAPHTPEEQVPVEAEAPONIEDEAKEQIQLLHMHVHAHVHVEGEDLQOEDGPTGE 178
Db 61 VPPEEAPHTPEEQVPVEAEAPONIEDEAKEQIQLLHMHVHAHVHVEGEDLQOEDGPTGE 120

QY 179 PQOEDDEFLMATDVRDFFETLEPEVSHEETSHTSHVVEETVSQDCNQDMMEEMSEQENPDS 238
Db 121 PQOEDDEFLMATDVRDFFETLEPEVSHEETSHTSHVVEETVSQDCNQDMMEEMSEQENPDS 180

QY 239 SEPVEDERLHHTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
Db 181 SEPVEDERLHHTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 240

QY 299 SIFPVEEQEVPDET 313
Db 241 SIFPVEEQEVPDET 255

RESULT 10
US-09-849-626-1806
; Sequence 1806, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tongtong
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-849-626-1806

Query Match 32.8%; Score 1320; DB 9; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 59 MVIALLGWTSVAVVWFDLVYEEVLGKGIYDADGDFDVEDDAAKVLGLKERSTSEPA 118

```

```
Db      241 SIFFVEEQVEPPDT 255

RESULT 12
US-10-017-754-1806
; Sequence 1806, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongcong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Warnerakis, Margarita
; APPLICANT: Carter, Darick
; APPLICANT: Panger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C18
; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-754-1806

Query Match      32.8%; Score 1320; DB 14; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      59 MVIALLGWVTSVAVVWFDLVYEEVLGKGIYDADGDFDVGDDAKVLLGLKERS*SEPA 118
Db      1 MVIALLGWVTSVAVVWFDLVYEEVLGKGIYDADGDFDVGDDAKVLLGLKERS*SEPA 60
QY      119 VPPEAEAPHTPEEQVPVEAPQNIIEDEAKEQIQSLLEHVMHAEHVEGEDLQOEDGPTGE 178
Db      61 VPPEAEAPHTPEEQVPVEAPQNIIEDEAKEQIQSLLEHVMHAEHVEGEDLQOEDGPTGE 120
QY      179 PQOEDDEFMATDVRFFETLEPEVSHHEETSHYVETVSQCNQDMEMMSEQENPDS 238
Db      121 PQOEDDEFMATDVRFFETLEPEVSHHEETSHYVETVSQCNQDMEMMSEQENPDS 180
QY      239 SEPVEDERLHHTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
Db      181 SEPVEDERLHHTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 240
QY      299 SIFFVEEQVEPPDT 313
Db      241 SIFFVEEQVEPPDT 255

US-10-017-754-1806

Query Match      32.8%; Score 1320; DB 14; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      59 MVIALLGWVTSVAVVWFDLVYEEVLGKGIYDADGDFDVGDDAKVLLGLKERS*SEPA 118
Db      1 MVIALLGWVTSVAVVWFDLVYEEVLGKGIYDADGDFDVGDDAKVLLGLKERS*SEPA 60
QY      119 VPPEAEAPHTPEEQVPVEAPQNIIEDEAKEQIQSLLEHVMHAEHVEGEDLQOEDGPTGE 178
Db      61 VPPEAEAPHTPEEQVPVEAPQNIIEDEAKEQIQSLLEHVMHAEHVEGEDLQOEDGPTGE 120
QY      179 PQOEDDEFMATDVRFFETLEPEVSHHEETSHYVETVSQCNQDMEMMSEQENPDS 238
Db      121 PQOEDDEFMATDVRFFETLEPEVSHHEETSHYVETVSQCNQDMEMMSEQENPDS 180
QY      239 SEPVEDERLHHTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
Db      181 SEPVEDERLHHTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 240
QY      299 SIFFVEEQVEPPDT 313
Db      241 SIFFVEEQVEPPDT 255

US-10-113-872-1806
; Sequence 1806, Application US/10113872
; Publication No. US20030170255A1
; GENERAL INFORMATION:
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Panger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C19
; CURRENT APPLICATION NUMBER: US/10/113,872

RESULT 13
US-10-113-872-1806
; Sequence 1806, Application US/10113872
; Publication No. US20030170255A1
; GENERAL INFORMATION:
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Panger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C19
; CURRENT APPLICATION NUMBER: US/10/113,872

Query Match      6.8%; Score 275; DB 14; Length 250;
Best Local Similarity 33.3%; Pred. No. 9.2e-12;
Matches 65; Conservative 31; Mismatches 91; Indels 8; Gaps 4;

QY      569 GLKAQPMWTPKGTGYTEL---VKSLEARNWKLIRDEGLAVMDKAKGLFLPSDENLRKGDW 625
Db      48 GLSPTPWHPD--YAYAEAPVHVELEASHQAIIKEELKVAWSAREAPSDYEHVLTRODNW 105
QY      626 SQFTLMQOGRNENACKGAPKTCITLLEKPPETTG--CRRGQIKYISIMHPGTHVWPHPTGPT 683
Db      106 QALYLFREGGITESAATVPIAYQVLKDVAVDTGKICPLECHFSTLLPGAVIEPHCDLW 165
```


